

Search time (sec): 1636.450000

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est1:AL513944	+	1156.00	905.12	3.6e-41	891	AL513944 AL513944 LTI_NFL006.P1
gb_est2:BF536373	+	1093.00	857.77	1.6e-38	744	BF536373 602051728F1 NCL_CGAP.S
gb_est1:BF808908	+	904.00	714.24	1.5e-30	517	BF808908 214215 MARC 2B0V Bos t
gb_est1:AW230557	+	850.00	672.72	3.2e-28	495	AW230557 u065all.y1 NCL_CGAP.Ma
gb_est2:BT333032	+	848.50	666.10	6.8e-28	866	BT333032 602983024F1 NCL_CGAP.I
gb_est2:BX31602	+	828.00	654.83	3.4e-27	592	BX31602 UT-R-DX0-ckr-1-07-0-UT
gb_est2:BX31602	+	828.00	654.83	3.4e-27	592	BX31602 UT-R-DX0-ckr-1-07-0-UT
gb_est2:BT094238	+	808.50	635.78	3.6e-26	865	BT094238 602860118F1 NIH_MGC.10
gb_est2:BT094238	+	808.50	635.78	3.6e-26	865	BT094238 602860118F1 NIH_MGC.10
gb_est1:AW763191	+	785.00	606.47	1.0e-24	513	AW763191 uc60h06.y1 NCL_CGAP.Ma
gb_est2:BM148104	+	737.00	586.11	2.1e-23	437	BM148104 TCAAP048352 Pediatr.ic
gb_est1:AF683106	+	729.00	577.26	6.6e-23	597	AF683106 t001d11.x1 NCL_CGAP.UB
gb_est2:BG432423	+	681.00	539.63	8.2e-21	625	BG432423 60249152F1 NIH_MGC.18
gb_est2:BF993048	+	625.00	500.87	1.2e-18	360	BF993048 IL5-GN0178-311000-204
gb_est2:BF993048	+	625.00	500.87	1.2e-18	360	BF993048 IL5-GN0178-311000-204
gb_est2:BF752541	+	606.00	482.21	1.3e-17	571	BF752541 PM1-GN0178-311000-204
gb_est2:BF752541	+	606.00	482.21	1.3e-17	571	BF752541 PM1-GN0178-311000-204
gb_est2:BG005397	+	594.00	476.59	2.7e-17	370	BG005397 PM1-GN0180-241100-003
gb_est2:BG005397	+	594.00	476.59	2.7e-17	370	BG005397 PM1-GN0180-241100-003
gb_est2:BG090891	+	579.00	463.31	1.5e-16	449	BG090891 HOAY-1-H11 HOA (Human
gb_est2:BF993027	+	551.00	444.36	1.7e-15	324	BF993027 IL5-GN0178-311000-204
gb_est2:BF993027	+	551.00	444.36	1.7e-15	324	BF993027 IL5-GN0178-311000-204
gb_est2:BF383613	-	543.00	431.48	8.7e-15	771	BF383613 Q0308395F1 NIH_MGC.12
gb_est2:BF987707	-	535.00	430.75	9.6e-15	373	BF987707 G0308395F1 NIH_MGC.12
gb_est2:BF987707	-	535.00	430.75	9.6e-15	373	BF987707 G0308395F1 NIH_MGC.12
gb_est1:AL637480	-	534.00	431.57	8.6e-15	309	AL637480 t070707.x1 NCL_CGAP.GC
gb_est2:DF8761	-	532.00	430.68	9.6e-15	286	DF8761 HUM510H03B Human placen
gb_est2:BG008790	+	529.50	426.44	1.7e-14	375	BG008790 MR3-GN0180-241100-013
gb_est2:BF989747	+	503.00	407.58	1.9e-13	373	BF989747 IL5-GN0178-311000-195
gb_est1:BF855583	+	486.00	390.28	1.7e-12	498	BF855583 BB855583 RIKEN full-1e
gb_est2:BF986444	+	480.00	385.47	3.2e-12	507	BF986444 QVA-GN0143-031000-445
gb_est2:BF986444	+	480.00	385.47	3.2e-12	507	BF986444 QVA-GN0143-031000-445
gb_est1:AU140343	+	474.00	376.80	9.7e-12	814	AU140343 AU140343 PLACE2 Homo s
gb_est2:BF989357	-	461.00	376.34	1.0e-11	262	BF989357 IL5-GN0179-201000-185
gb_est2:BF989357	-	461.00	376.34	1.0e-11	262	BF989357 IL5-GN0179-201000-185
gb_est2:BF645713	+	450.00	366.35	3.7e-11	375	BF645713 G03275026F1 NIH_CGAP.M
gb_est2:BF645713	+	450.00	366.35	3.7e-11	375	BF645713 G03275026F1 NIH_CGAP.M
gb_est1:AL665815	+	445.00	352.43	3.2e-10	1039	AL665815 AL665815 directional
gb_est2:BP443732	+	442.00	356.28	1.3e-10	479	BP443732 261436 MARC 2P1G Sus s
gb_est2:BP443732	+	442.00	356.28	1.3e-10	479	BP443732 261436 MARC 2P1G Sus s
gb_est1:AL617554	+	435.50	348.16	3.8e-10	704	AL617554 xn55903.x1 Soares_NICE
gb_est2:BM150365	+	426.00	349.06	3.4e-10	266	BM150365 TCAAP09028F Pediatr

```

67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84
421 GAGCCGAGATCTAGCTCCCTACAGTGCAGCTCCACCCGAGTCTKCHCC 470
84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100
471 TCCTCCCGGGCCCAAGACATCCAGAGCCCTCTCCCGGGGGCCAG 520
101 GlyProArgProGlnThrSerProGlnThrLeuProLeuTyArgThrG 117
521 GTTCCAGACCCAGACTTCTCCAGAACCTTCCCTTGTACAGGACACA 570
117 nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyA 134
571 GTCCTGGGAAGGGTGGCCACTTCGAGTCCGCTCCACCTAGGGA 620
134 rGluGluThrGlnGluLeuArgAlaAlaArgArgSerArgLeuArgasp 150
621 GAGAGGAGACCCAGGAGATTCAGCGGGCCAGGAGTCCCGGCTTCGAG 670
151 ProLeuLysProGlyMetPheGlyTyArgValProPheAlaLeuPr 167
671 CCCATCAGCCAGGATGTCGTTATGGAGAGTCCCTTTGCAATTGCC 720
167 oLeuHisArgAsnArgArgHisProArgSerProProArgSerGluLeu 184
721 ACTGCACCGGAACCCAGCAACCTCGGAGCCACCCAGATCTGAGCTGT 770
184 eLeuLeuSerSerArgGlyGluAlaLeuProSerProThrProArg 200
771 CCTGATCTCTCTAGAGGGAGAGCCTATTCGCTCCCTACTCCAGA 820
201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProth 217
821 GCAGAGCCATTCGCGCAACGGCAGCCGCCCAACTGAGCTCCCTCCAC 870
217 rGluLeuSerValHis 222
871 AGAAGTGTCTGTCCAC 886

seq_name: gb_est2.BF536373
seq_documentation_block:
LOCUS BF536373 744 bp mRNA linear EST 11-DEC-2000
DEFINITION 602051728F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4191007 5',
mRNA sequence.
ACCESSION BF536373
VERSION BF536373.1 GI:11623741
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 744)
NTH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9519 row: g column: 08
High quality sequence stop: 685.
Location/Qualifiers
1..744
/organism="Mus musculus"
/strain="FVB/N"

```

```

/db_xref="taxon:10090"
/clone_image="4191007"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 139 a 231 c 208 g 166 t
ORIGIN
alignment_scores:
Quality: 1093.00 Length: 259
Ratio: 4.836 Gaps: 2
Percent Similarity: 87.259 Percent Identity: 79.151
alignment_block:
US-10-041-770-2 x BF536373 ..
Align seg 1/1 to: BF536373 from: 1 to: 744
428 ArgPheTyValArgHisThrGluLysValGlnAspGlyThrLeuCysG 444
2 CGATTCTATGTCCGACACACTGAAAGGTGCAGATGGAACCTGTGTCA 51
444 nProGlyAlaProAspLeuCysValAlaGlyArgCysLeuSer.ProGly 460
52 GCCTGGATCCCTAGACATCTGTGGTGGACGCTGCTGCGCTCCG... 98
461 CysAspGlyLeuLeuGlySerGlyArgArgProAspGlyCysGlyValCy 477
99 .....GATGGCTGGGAGTCTG 115
477 sGlyGlyAspSerThrCysArgLeuValSerGlyAsnLeuThrAspA 494
116 CGGGGTGATGGTTCTACCTGCCGCTGGTTTCGGAAATCTCACTGACC 165
494 rGlyGlyProLeuGlyTyGlnLysIleLeuTrpIleProAlaGlyAla 510
166 GAGGGGGCCCTTGGCTATCAGAGATCTTGGTATCCCTGCGGGGGCC 215
511 LeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyLeuAlaLe 527
216 TCCACCTTCACATTCCAGTTCGCCAGTCCAGTCCCAATTACCTCGCACT 265
527 uArgGlyProGlyArgSerIleIleAsnGlyAsnTrpAlaValasp 544
266 CCGAGGGCCCTGGGGCCGCTCCATTATCAATGGGAAGTGGCTGTGGATC 315
544 roProGlySerTyArgAlaGlyGlyThrValPheArgTyArgAsnArgPro 560
316 CTCAGGGTCTATACAGCCATTGGGACTGTCTTCCAGTATACCGTCTCT 365
561 ProArgGluGluGlyLysGlyGlySerLeuSerAlaGluGlyProThr 577
366 CCACGGGAAGCAAGCAAGGGGAGAGTCTGTACAGAGAAGGCCCTACCAC 415
577 rGlnProValAspValTyMetIlePheGlnGluGluAsnProGlyValp 594
416 TCAGCTGTGGAGCTCTATATGATCTTTCAGAGAGACACCCAGGTGTTT 465
594 heTyGlnTyValIleSerSerProProPheIleLeuGluAsnProThr 610
466 TTTATCAGTATGTCTATCTCTTCCCTCTCTGAGTCTAGAGAGTCTCTCC 515
611 ProGluProProValProGlnLeuGlnProGluLeuArgValGluPr 627
516 ACAAGCCCTCCAGCCCTTACCCCTCAGCTCAGATCTGAGGGGGGAGCC 565
627 oProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnA 644
566 CCTACTCCCTCAGCCCCCCCCCGGAGTCCGGGACACAGGACCCCTCCAC 615

```

644 rgglValArgIleProGlnMetProAlaProHisProArgThrPro 660  
|||||  
616 GTGAGTGGGATCCCAAGTCCCTCCGACTCGTGTGAGGACGCC 665  
|||||  
661 LeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSe 677  
:::|||||  
666 ATGGGGTCTCTCTGCTGGATACCTGGGAACAGTGGGGCACTCTGAGTGTTC 715  
|||||

seq\_name: gb\_est2:BE808908

seq\_documentation\_block:  
LOCUS BE808908 517 bp mRNA linear EST 25-APR-2001  
DEFINITION 214215 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE808908  
VERSION BE808908.1 GI:10240020  
KEYWORDS EST.

ORGANISM

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE

1 (bases 1 to 517)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
,G.L., Heaton,M.P., Laegreid,W., Rohrer,M.G.A., Chitko-McKown,C.G.,  
Partee,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle

JOURNAL

Genome Res. 11 (4), 626-630 (2001)

MEDLINE

21180013

COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 70 row: N column: 6

Seq primer: ATTTAGTGACACTATAG.

FEATURES

source

1..517

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 2BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="vector: pcwv sport6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."

BASE COUNT 84 a 175 c 179 g 79 t

ORIGIN

alignment\_scores:

Quality: 904.00 Length: 167

Ratio: 5.722 Gaps: 0

Percent Similarity: 94.611 Percent Identity: 93.413

alignment\_block:

US-10-041-770-2 x BE808908

Align seg 1/1 to: BE808908 from: 1 to: 517

691 CysIleSerArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAl 707  
|||||  
1 TGCATTCTCGAGATCAGGAGAGGAGCTGGATGAACGACGCTGTGCCAT 50  
|||||  
707 aGlyAlaArgProAlaSerProGluProCysHisGlyThrProCysP 724  
|||||  
51 GGGTGCCAGGCCCCAGCCTCCAGGAGCCTGCCACGCCCCCGGTGCC 100  
|||||  
724 roProTyrTrpGluAlaGlyclufTrpSerCysSerArgSerCysGly 740  
|||||  
101 CACCATACTGGAGCGCGGAGTGGAGCTCTGCACGCCGTTCGTGTGGA 150  
|||||  
741 ProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757  
|||||  
151 CCCGGCACCCAGCAGCTCAGCTACGTGCGGCGGAGGAGTTGGGGTGG 200  
|||||  
757 yGlySerSerValProGluArgCysGlyHisLeuProArgProAsnI 774  
|||||  
201 CGGCTCCTCAGTGCCTAGAGCGCTGGGGCACCTGCCCGACCCCAACA 250  
|||||  
774 leThrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySer 790  
|||||  
251 TCACCCAGCCCTGCCAGCTGGCCTCTGTGCCCATTTGGAGTTTCGTCA 300  
|||||  
791 ProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerArgG 807  
|||||  
301 CCCGTGGAGTCACTCTCTGTGCGATGGGGCGCGTCAGAGGACCGGCA 350  
|||||  
807 nValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCys 824  
|||||  
351 GGTCCGTGTGTGCGCAACAAACAGGGGATGAAGTGAAGTGAAGTGAAG 400  
|||||  
824 laSerGlyProProGlnProProSerArgGluAlaCysAspMetGlyPro 840  
|||||  
401 CCTCATGGCCCCCGCCCCCAGCAGAGAGGCTTGGACATGGGGCCC 450  
|||||  
841 CysThrThrAlaTrpPheHisSerAspTrpSerSerLysValSerProG 857  
|||||  
451 TGCACACGGCGCTGTTCCACAGCGACTGGAGCTCCAAGTCTCAGCGGA 500  
|||||

857 u 857

501 g 501

seq\_name: gb\_est1:AW230557

seq\_documentation\_block:

LOCUS AW230557 495 bp mRNA linear EST 10-DEC-1999  
DEFINITION uc65a11.y1 NCI\_CGAP\_Maml Mus musculus cDNA clone IMAGE:2647388 5'  
similar to TR:060345 O60345 KIAA0605 PROTEIN.; mRNA sequence.

ACCESSION AW230557

VERSION AW230557.1 GI:6559853

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 495)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

MG1:1027840

Seq primer: -40RP from Gibco  
High quality sequence stop: 357.

#### FEATURES

source  
1..495  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2647388"  
/clone\_lib="NCI\_CGAP Mam1"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="10 months, virgin"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Sall;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 84 a 155 c 149 g 107 t  
ORIGIN

#### alignment\_scores:

Quality: 850.00 Length: 164  
Ratio: 5.484 Gaps: 0  
Percent Similarity: 94.512 Percent Identity: 87.805

#### alignment\_block:

US-10-041-770-2 x AW230557 ..  
Align seg 1/1 to: AW230557 from: 1 to: 495  
647 ArgIleProGlnMetProAlaProHisProArgThrProLeuGlyse 663  
1 CGGATTCCTCCCAAGTGCCTCTCGAGCTGTCGAGCAGCAGCCATGGGTC 50  
663 rProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSerC 580  
51 TTTCTGTGGATCTGAAACAAGTGGGCACTGTGAGTTTTCAGCATCT 100  
680 ySGLYLYSGLYValTrpArgProIlePheLeuCysIleSerArgGluSer 696  
101 GTGCAAGGTGTTTGGACCCCATTTCTCTGTCATTTCCGTCAGTCA 150  
697 GlyGluGluLeuaspGluArgSerCysAlaAlaGlyAlaArgProAla 713  
151 GGAGAGGAGTTGGATGAACAGAGCTGTGTGGGGCCAGACCCCCAGC 200  
713 aSerProGluProCysHisGlyThrProCysProProTyrTrpGluAla 730  
201 TTCCCTGTGAACCTGCCATGGACCCCGTCTCTCATCTGGAGGCTG 250  
730 lyGluTrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArg 746  
251 GCAGTGGAGATCTCTGCAGCCGATCTGTGGCCCTGGCACCCAGCCGC 300  
747 GlnLeuGlnCysArgGlnGluPheGlyGlyGlySerSerValProPr 763  
301 CAGCTACTCTGCAGACAGAGATTCCGAGGTGGTGGCTCTCGGTACTCC 350  
763 oGluArgCysGlyHisLeuProArgProHisLeuThrGlnSerCysGln 780  
351 AGAGCGCTGTGGACATCTCCCGCGGCCAACATCACCCAGCCTTGTG 400  
780 euArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSer 796  
401 TGACCTCTGTGGCCACTGGGAGATTAGTCCCTGGAGCCAGTGTCT 450  
797 ValArgCysGlyArgGlyGlnArgSerArgGlnValArgCys 810  
451 GTCCGT 492

seq\_name: gb\_est2:BI330332

#### seq\_documentation\_block:

BI330332 866 bp mRNA linear EST 30-JUL-2001  
LOCUS 602983024F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5135860 5',  
DEFINITION mRNA sequence.

ACCESSION BI330332  
VERSION BI330332.1 GI:15014989

#### KEYWORDS

SOURCE house mouse.

#### ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

#### REFERENCE

1 (bases 1 to 866)  
NIH-MGC http://mgc.nci.nih.gov/.

#### AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

#### TITLE

Unpublished (1999)

#### JOURNAL

#### COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM11333 row: p column: 05  
High quality sequence start: 3  
High quality sequence stop: 548.

#### FEATURES

##### source

1..866  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5135860"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: Sall; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 149 a 332 c 237 g 148 t

#### ORIGIN

#### alignment\_scores:

Quality: 848.50 Length: 286  
Ratio: 3.839 Gaps: 13  
Percent Similarity: 77.273 Percent Identity: 67.483

#### alignment\_block:

US-10-041-770-2 x BI330332 ..  
Align seg 1/1 to: BI330332 from: 1 to: 866  
376 GlnAlaProCysProGluGlnProAspProArgAlaLeuGlnCysAl 392  
2 CAGGAGCCTTGCCCCCTGGAGATGCCAGACCCAGGCCCTACAGTGTGC 51  
392 aAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrGlnTrpGluProp 409  
52 CCCTTTTGACTCCAGGAATTCATGGCCAGCTGTACCAGTGGAGCCCT 101  
409 heThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProArg 425  
102 TCACCGAGTTTCAGGGCTCCAGCCCTGTGACTGACTGACCTGCCGCCCT 151  
426 GlyPheArgPheTyrValArgHisThrGluLysValGlnAspGlyThrLe 442  
152 GCGTTCGCGATTCTATGTCGACACACTGAAAGGTGCAGATGGAACCT 201  
442 uCysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSerP 459



451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
Oligo-dt track not found, Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=No. Location/Qualifiers

FEATURES  
source

1. .592  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DY0-ckr-1-07-0-UI"  
/clone\_lib="UI-R-DY0"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not 1; Site 2: Eco RI; UI-R-DY0 is a  
non-normalized Rat cartilage library (RC) constructed in  
pT37 PAC vector according to the procedure described by  
Bonaldi, Lennon & Soares (Genome Research Genome 6:  
791-806, 1996). The oligonucleotide used to prime first  
strand synthesis contained the sequence tag CTAATGGACG  
between the Not I cloning site and d18 stretch. The Rat  
cartilage tissue was provided by Dr Jeff Stevens at the  
University of Iowa.  
TAG\_SEQ=None found"

BASE COUNT 110 a 181 c 192 g 108 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 828.00 Length: 160  
Ratio: 5.483 Gaps: 1  
Percent Similarity: 94.375 Percent Identity: 88.750

alignment\_block:  
US-10-041-770-2 x BM391602 ..  
Align seg 1/1 to: BM391602 from: 1 to: 592

699 GluLeuAspGluArgSerCysAlaAla GlyAlaArgProAlaSerP 715  
27 CAGTTGGATGAACAGAGCTGTCCTGGCGGGGCGGACCCAGCTGCC 76  
715 rogluProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGlu 731  
77 CTGAATCTGTCATAGACACCGTGTCTCCCATCTGGAGGCTGGCGAG 126  
732 TrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLe 748  
127 TGGAGCTCTGTCAGCGCTCTCTGTGGCGCTGGCACCAGCATCGCCAGCT 176  
748 uGlnCysArgGlnGluPheGlyGlyGlySerSerValProGluA 765  
177 GCTCTGCACAGAGGAATTCGGAGGCGGTGGCTCTCAGTGCCTCCAGAGC 226  
765 rGlyGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArg 781  
227 GTTGGGACATCTCCCGCGGCCAACATCACCCAGTCTCTGTCAGCTGGC 276  
782 LeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSerValar 798  
277 CTTTGTGGCACTGGAGATTAAGTCTCCCGCTGGAGCCAGTCTCTGTGCG 326  
798 gCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnG 815  
327 CAGTGTCTGTGTGTCAGAGGAGCGGCAAGTGGGTGTGTGTGAAGTAATG 376  
815 lyAspGluValSerGlnGluCysAlaSerGlyProProGlnProPro 831

202 GTGTACGCTGATCCCTAGACATCTGTGTGCTGGACGCTCCTCGAGCC 251  
459 rogluCysAspGlyLeuLeuGlySerGlyArgPro AspGlyCysG1 475  
252 CTGGCTGTGATGGGTCTCTGTGCTGGAGCGCTCCGCGATGGTGGCG 301  
475 yValCysGlyGlyAsp AspSerThrCysArgLeuValSerGly AsnLe 491  
302 AGTCTGGGGGTGACTGTCTTACCTGCGGCTGGTTCGGGCAAACT 351  
491 uThrAspArgGlyGlyProLeuGly TyrGln LysIleLeuTrpIlePr 507  
352 CACTGACCGAGGGGCGCTTGGGCTATCAGATGATCTGTGGATCCC 401  
507 oAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTr 524  
402 TGGCGGGGCTCCACCTTCACATTCCTCCAGTGGAGCAGT.TCCAAT 450  
524 yLeuAlaLeuArgGlyProGlyArgSerIle.IleAsnGlyAsnTr 540  
451 ACCTGCACTCGAGGCGCTGGGGCGCTCCATTCATCATCTGTTGGAAC 500  
540 pAlaValAspProPro GlySerThrArgAlaGlyGlyThrValPheArg 556  
501 GGTGTGGATCTCCAGGCTCTCTACAGCCATCCGACTGTCTTCCAG 550  
557 TyrAsnArgProProArg GluGlu...GlyLysGlyGluSerLeuSerA 572  
551 TATAAGCTCTCCACGGGCAAGACAGGCGCGGAGGAGTCTGTCTCAG 600  
572 laGluGlyProThrThrGlnProVal AspValTyrMetIlePheGlnI 588  
601 CCAAGGCTTACCACTACGCTGCGCGCTCTCTCTGACCCACCCAC 650  
588 uGluAsnProGlyValPheTyrGlnTyrValIleSerSerProProI 605  
651 CGAGGACCCCGGAGGTGTCGCGGACCCGCGGATCTATCCACCCCG 700  
605 leLeuGluAsnProThrProGluPro..... 613  
701 TCCCGGCGAGCGGAGGAGGCGCTCTCACAGACGCTCCGACCCAT 750  
614 .....ProValProGlnLeuGlnProGluIleLeuArgValGluProPr 628  
751 GTAGCCCCCAGGCGCGAGTGTG...CCGAGGGGCGCAACCAACAGAGCC 797  
628 oLeuAlaProAlaProArgProAlaArgThrProGlyThr 641  
798 CACAGCCCGCGGCGGCGGCGGCGG...ACACGACCAAG 834

seq\_name: gb\_est2:BM391602  
seq\_documentation\_block:  
LOCUS BM391602 592 bp mRNA linear EST 17-JAN-2002  
DEFINITION UI-R-DY0-ckr-1-07-0-UI.s1 UI-R-DY0 Rattus norvegicus cDNA clone  
UI-R-DY0-ckr-1-07-0-UI 3', mRNA sequence.  
ACCESSION BM391602  
VERSION BM391602.1 GI:18191655  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE (bases 1 to 592)  
AUTHORS Bonaldi,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477  
COMMENT Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa

```
377 GCCATGAGTGGGCAAGCAGGAGTGTCTCGGGCCCCCGCCACTCCC 426
832 SerArgGluAlaCysAspMetGlyProCysThrThrAlaTIPheHisSe 848
427 AGCAGAGAGCCCTGTGACATGGCCCTGTACACAGCCTGTTCTACAG 476
848 rAspTIPserSerLysValserProGlu 857
477 TGACTGAGTTCGAAGTGTCTCAGCCGAG 504
```

```
seq_name: gb_est2:BI094238
seq_documentation_block:
LOCUS      BI094238               865 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION 602860118F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5001173 5',
            mRNA sequence.
ACCESSION  BI094238
VERSION    BI094238.1 GI:14512568
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 865)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LRAM11034 row: d column: 06
            High quality sequence stop: 510.
```

```
FEATURES
            Location/Qualifiers
                1..865
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5001173"
                    /clone_lib="NIH_MGC_10"
                    /cell_line="MGC36"
                    /lab_host="DH10B"
                    /note="org: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                    Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
                    Average insert size 1.5 kb. Library prepared by Life
                    Technologies."
BASE COUNT  240 a 187 c 303 g 135 t
ORIGIN
```

```
alignment_scores:
    Quality: 808.50      Length: 204
    Ratio: 4.620        Gaps: 6
    Percent Similarity: 85.784      Percent Identity: 82.843
alignment_block:
US-10-041-770-2 x BI094238 ..
Align seg 1/1 to: BI094238 from: 1 to: 865
425 ArgGlyPheArgPheThrValArgHisThrGluLysValGlnAspGlyTh 441
1 CGTGGCTTCGGCTCTATGTCCGCTCACACTGAAAGGTCAGGATGGGAC 50
441 rLeuGlyGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeus 458
51 CCTGTGTGAGGCTGGAGCCCTGCATCTGTGTGGCTGGAGAGCTGTCTGA 100
```

```
458 erProGlyCysAspGlyIleLeuGlySerGlyArgArgProAspGly_Cy 474
101 GCCCGGGCTGTGATGGGATCCTTGGCTCTGGCAGGGCGTCTGATGGCTT 150
474 sglyValCysglyGlyAspSerThrCysArgLeuValserGlyAsnL 491
151 TGGAGTCTGTGGGGGTGATGATCTACCTGTGCGCTTGTTCGGGAACC 200
491 euThrAspArgGlyGlyProLeuGlyTyrGlnLysIleLeuThrPilePro 507
201 TCACTGACCGAGGGGGCCCTGGGCTATCAGAAGATCTTGTGATTCCA 250
508 alaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTy 524
251 GGGGGAGCCCTTGGGCTCCAGATTGCCAGTCCGGCTAGCTCCAACTA 300
524 rleuAlaLeuArgGlyProGlyGlyArg.SerIleIleAsnGlyAsn... 539
301 CCGTGCACCTTGTGGCCCTGGGGCTGTGCTCATCATCAATGGGAACCTG 350
540 rtpAlaValAspProProGlySerTyrArgAlaGlyGlyThrValPheAr 556
351 TGGCTTGTGGATCCCGCTGGGTCTCTACAGGCGGGGACCGTCTTCG 400
556 gtyrAsnArgProProArgGlu.GluGlyLysGlyGluSerLeuSerAla 572
401 ATATAACCGTCTCTCCAGGAAGGAGGCAAAAGGGAGAGTCTGTGCGCT 450
573 GluGlyProThrThrGlnProValAspValTyrMetIlePheGlnGluG1 589
451 GAAGGCCCAACAC.CAGCTGTGGATGTCTATATGATCTTTCAGGAGA 499
589 uAsnProGlyValPheThrGlnTyrValIleSerSerProProPile. 605
500 AAACCCAGCGGTAGATAATCGAGTAGTCTGCTGAGTCTGTGACGCGTCG 549
606 LeuGluAsnPro.....ThrProGluProProValPr 616
550 CTCGCAATCTTGGAGAAAGCGCAAGACACAGAGCGCGCCCTAGAGC 599
616 ogln 617
600 GCAC 603
```

```
seq_name: gb_est1:AW763191
seq_documentation_block:
LOCUS      AW763191               513 bp      mRNA      linear      EST 04-MAY-2000
DEFINITION ur-60h06.y1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:3154715 5',
            similar to WP:F25H8.3 CE05729 THROMBOSPONDIN LIKE ;, mRNA sequence.
ACCESSION  AW763191
VERSION    AW763191.1 GI:7695127
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 513)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
MGI:1057471
```

seq primer: -40RP from Gibco  
High quality sequence stop: 336.

FEATURES  
source  
1. .513  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="129 - C57/B6 - FVBN"  
/db\_xref="taxon:10090"  
/clone\_image="3134715"  
/clone\_lib="NCI\_CGAP\_Mam3"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="10 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: oligo dT. Library constructed by life technologies. Investigators providing samples: lothar Hennighausen/Chu-xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

BASE COUNT 98 a 159 c 161 g 95 t  
ORIGIN

alignment\_scores:  
Quality: 765.00 Length: 146  
Ratio: 5.543 Gaps: 0  
Percent Similarity: 94.521 Percent Identity: 87.671

alignment\_block:  
US-10-041-770-2 x AW763191 ..

Align seg 1/1 to: AW763191 from: 1 to: 513

708 GLYALAArgProAlaSerProGluProCysHisGlyThrProCysPr 724  
7 GGGGCCAGACCCCGAGCTTCCCTCGAACCTGGACCGCCCGTGTCC 56  
724 oProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyP 741  
57 TCCATACTGGAGGCTGGCAGTGGACATCTCTGCAGCCGATCTGTGGCC 106  
741 roGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757  
107 CTGGACCCAGCAGCCGACGCTCTCTGCAGACGAGATTCGAGGTGT 156  
758 GlySerSerValProGluArgCysGlyHisLeuProArgProAsn11 774  
157 GCGTCCTCGTACTCTCAGAGCGCTGTGGACATCTCCCGGCCCAACAT 206  
774 eThrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySerP 791  
207 CACCCAGCCCTGTGACGTGACCTCTGTGCCACTGGAGATTAGCTCC 256  
791 roTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerArgGln 807  
257 CTGGAGCCAGTGTCTGTGCGCTGTGGTGTGTGAGGAGCCGCGAA 306  
808 ValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAl 824  
307 GTTCGGTGTGTGGCAGTAATGTGTGATGAGTGGACAAAGCAGGAGTGC 356  
824 aSerGlyProProGlnProProSerArgGluAlaCysAspMetGlyProC 841  
357 TTCAGGCCCGCCCGCCCGCAGCAAGAGGCTGTGACATGGGCCCT 406  
841 ysthThrAlaTrpPheHisSerAspTrpSerSerLys 853  
407 GCACCAACACCTGTTCTACAAGGACGTGAGTTGCAAG 444

seq\_name: gb\_est2:BM148104

seq\_documentation\_block:  
LOCUS BM148104 437 bp mRNA linear EST 30-NOV-2001  
DEFINITION TCAAP108352 Pediatric acute myelogenous leukemia cell (FAB M1)

Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP8352, mRNA sequence.  
BM148104  
BM148104.1 GI:17168519  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
Wei,Y., Tsang,Y.T.M., Mei,G., Ku,J.M., Ali-Osman,F.R. Jr., Gunaratne,P.H., Muzny,D., Bouck,J., Gibbs,R.A. and Margolin,J.F. Pediatric Leukemia cDNA Sequencing Project (2001)  
Unpublished (2001)  
Contact: Dr. Judith F. Margolin  
Texas Children's Cancer Center and Human Genome Sequencing Center at Baylor College of Medicine  
1102 Bates, MC3-3320 Houston, TX 77030, USA  
Tel: 832-824-4536  
Fax: 832-825-4038  
Email: clones@tccc.org  
Seq primer: M13 primer  
Location/Qualifiers  
1. .437  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TCAAP8352"  
/clone\_lib="Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project-TCAA"  
/sex="male"  
/tissue\_type="leukopheresis"  
/cell\_type="myeloid cell"  
/dev\_stage="pediatric 6 years"  
/lab\_host="DH10B"  
/note="Vector: lambda pSB; Site.1: BamHI; Site.2: EcoRI; First strand cDNA was primed with an anchored XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCCGAGGAGGAG(T)VN 3'; V=A,C,G; N=A,C,G,T] and then dG tailed. Second strand was primed with a BamHI-dC primer [5'AGAGAGCTCGATCGCGCCGCAATAATAAT(C) 3']. Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda pSB vector. Library went through one round of normalization. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoxa S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb. 28, 1997)"

BASE COUNT 76 a 151 c 131 g 79 t  
ORIGIN

alignment\_scores:  
Quality: 737.00 Length: 139  
Ratio: 5.459 Gaps: 0  
Percent Similarity: 97.122 Percent Identity: 96.403

alignment\_block:  
US-10-041-770-2 x BM148104 ..

Align seg 1/1 to: BM148104 from: 1 to: 437

7 ArgProTrpLeuTyrLeuLeuLeuLeuSerLeuProGlnLeuCysLe 23  
16 AGGCCCTGGGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 65  
23 uAspGlnGluValLeuSerGlyHisSerLeuGlnThrProThrGluGluG 40  
66 GCATCAGAGAGGTGTGTCGGGACACTCTCTTCAGACACCTACAGAGAGG 115  
40 lyGlnGlyProGluGlyValTrpGlyProTrpValGlnTrpAlaSerCys 56

116 GCCAGGGCCCCGAAGGTCTCTGGGACCTTGGGTCCAGTGGGCCCTCTTCG 165  
 57 SerGlnProCysGlyValGlyValGlnArgArgSerArgThrCysGlnLe 73  
 166 TCCAGCCCTCGGGGTGGGGTGCAGCGCAGGAGCCGACATGTACACT 215  
 73 uProThrValGlnLeuHisProSerLeuProLeuProProProProPro 90  
 216 CCTACAGTGCAGCTCCACCGGAGTCTGCCCTCCCTCCCGGGCCCCCAA 265  
 90 rGHisProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThr 106  
 266 GACATCCAGAACCTCTCTCCCGGGGCCAGGGTCCAGACCCGAGACT 315  
 107 SerProGluThrLeuProLeuTyrrzrGlnSerArgGlyArgGlyG1 123  
 316 TCTCCAGAAACCTCCCTCTGTACAGGACACAGTCTCGGGGAAGGGTGG 365  
 123 yProLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGlu 140  
 366 CCACCTTCAAGGACCGCTGTCCACCTAGGGAGAGGAGAGCCAGGAGA 415  
 140 leArgAlaAlaArgArg 145  
 416 TTCCAGCGGACAGCAGG 432

seq\_name: gb\_est1:AI683106

seq\_documentation\_block:

LOCUS AI683106 597 bp mRNA linear EST 16-DEC-1999  
 DEFINITION tx01d11.x1 NCI\_CGAP\_Ut4 Homo sapiens cDNA clone IMAGE:2267925 3',  
 similar to TR:002661 002661 SCO-SPONDIN ; contains MSRI.t2 MSRI  
 repetitive element ; , mRNA sequence.

ACCESSION AI683106

VERSION AI683106.1 GI:4893278

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 597) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1132 Std Error: 0.00

Seq primer: -40MP from Gibco

High quality sequence stop: 369.

Location/Qualifiers

1..597

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2267925"

/clone\_lib="NCI\_CGAP\_Ut4"

/tissue\_type="serous papillary carcinoma, high grade, 2

pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: SalI;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.48 kb. Life Technologies catalog #:

11542-016"

BASE COUNT 118 a 162 c 180 g 136 t 1 others

ORIGIN

alignment\_scores:

Quality: 729.00 Length: 175

Ratio: 4.796 Gaps: 0

Percent Similarity: 86.857 Percent Identity: 78.286

alignment\_block:

US-10-041-770-2 x AI683106/rev ..

Align seg 1/1 to reverse of: AI683106 from: 1 to: 597

45 GlyValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysG1 61  
 597 GGTGTTTGGGGCCCTTGGGTCAAGTGGGCTCTTGTCTCCAGAGCCCTCGG 548  
 61 yValGlyValGlnArgArgSerArgThrCysGlnLeuProThrValGlnL 78  
 547 GGTGGGGTGCAGCGCAGGAGCGCAACATGTACAGTCTCTTACAGTGGAGC 498  
 78 euHisProSerLeuProLeuProProArgProProArgHisProGluAla 94  
 497 TCCACCGGAGTTTGGCCCTCCCTCCCGGGCCCCCAGACATCCAGAGCC 448  
 95 LeuLeuProArgGlyGlnGlyProArgProGlnThrSerProGluThrLe 111  
 447 TTCCTCCCGGGGCCAGGCTCCAGACCCAGATTTCCTCCAGAAACCT 398  
 111 uProLeuTyrrzrGlnSerArgGlyArgGlyGlyProLeuArgGlyP 128  
 397 CCCTTTTACAGGACCCAGATTTCGGGAAGGGTGGCCCTTTTCGGGGTC 348  
 128 roAlaSerHisLeuGlyArgGluGluThrGlnGluLeuArgAlaAlaArg 144  
 347 CCGTTCCCTTAGGGAGAGAGGAGCCAGGAGTTTCGAGCGGCCAGG 298  
 145 ArgSerArgLeuArgAspProIleLysProGlyMetPheGlyTyrGlyAr 161  
 297 GGGTCCCGGTTTCAGGCCCCCATCAAGCCAGGAATGTGGTTAATGGAG 248  
 161 gValProPheAlaLeuProLeuHisArgAsnArgHisProArgSerP 178  
 247 AGTGCCTTTCATGCCANTGCACGGGACCGCGGGCCCCCTGGGGCC 198  
 178 roProArgSerGluLeuSerLeuLeuSerSerArgGlyGluGluAlaLe 194  
 197 CCCCAGTTTGTAGCTGTCCCTGTTTTTTTAGGGGGGAGGGGTTTTT 148  
 195 ProSerProThrProArgAlaGluProPheSerAlaAsnGlySerProG1 211  
 147 CGTCCCTTACTCCAGAGCAGAGCCCTTTTCCGCAAGAGCGGCCCCCA 98  
 211 nThrGluLeuProProThrGluLeu 219  
 97 AATTGGGCTCCCTCCCGCAGAACTG 73

seq\_name: gb\_est2:BG438243

seq\_documentation\_block:

LOCUS BG438243

DEFINITION 602490152F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:462245 5',

mRNA sequence.

ACCESSION BG438243

VERSION BG438243.1 GI:13344749

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 625) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: crapbe-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP/Gazdar  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM386 row: g column: 14  
 High quality sequence stop: 406.

## FEATURES

source  
 1. .625  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:462245"  
 /clone\_lib="NIH\_MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pORF7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

BASE COUNT 188 a 161 c 174 g 102 t

## ORIGIN

alignment\_scores:  
 Quality: 681.00 Length: 180  
 Ratio: 4.540 Gaps: 6  
 Percent Similarity: 83.333 Percent Identity: 81.667

## alignment\_block:

US-10-041-770-2 x BG438243

Align seg 1/1 to: BG438243 from: 1 to: 625

473 GlyCysGlyValCysGlyGlyAspSerThrCysArgLeuValSer.G 489  
 |||||  
 3 GGCTGTGGAGTCGTGGGGTGATGATCTACCTGTCGCTGTTCTGG 52  
 489 lYAsnLeuThrAspArgGlyGlyProLeuGlyTyrClnLysIleLeuTyr 505  
 |||||  
 53 GGAACCTCACTGACCGAGGGGCCCTGGCTATCAGAAGATCTTGTGG 102  
 506 lLeProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSe 522  
 |||||  
 103 ATCCAGCGGGAGCCCTGGGCTCCAGATTGCCAGCTCGGCTAGCTC 152  
 522 rAsnTyrLeuAlaLeuArgGlyProGlyArgSerIleIleAsnGlyA 539  
 |||||  
 153 CAACCTACCTGGCCTCTGGGGCCCTGGGGCCGGTCCATCATCAATGGGA 202  
 539 snTrrAlaValAspProProGlySerTyrArgAlaGlyGlyThrValPhe 555  
 |||||  
 203 ACTGGGCTGTGGATCCCTGGGTCTCTACAGGCGCGGGACCGCTTT 252  
 556 ArgTyrAsnArgProProArg.GluGluGlyLysGlyGluSerLeuSerA 572  
 |||||  
 253 CGATATAACCGTCTCCCGAGTGAGGGGCAAGGGGAGAGTCTGTGG 302  
 572 lAgGlyProThrThrGlnProValAspValTyrMetIlePheGlnGlu 588  
 |||||  
 303 CTGAGAGCCCCACACCCAGCTGTGGATGCTATATGATCTTCAGAG 352  
 589 GluAsnProGlyValPheTyrGlnTyrValIleSerSerProProI 605  
 |||||  
 353 GAAACCCAGGCGT.TTTTATCAGTATGATCTCTTCACCTCTCCAAAT 401

605 eLeu...GluAsnPro.ThrProGlu...ProProValProGlnLeuGln 619  
 |||||  
 402 CTTTGAGGAAACCGACACGAGCGCGCTGTACCCACACAAGAT 451  
 |||||  
 620 ProGluIleLeuArgValGluProProLeuAlaProAlaProArgProAl 636  
 |||||  
 452 ACAGAA.....CGAGCAGA 465

636 aArgThrProGlyThrLeuGlnArgGlnValArg 647

466 GTCCTGGAGGAGCGACGACCAACGCCACACGCA 499

seq\_name: gb\_est2:BF993048

## seq\_documentation\_block:

LOCUS BF993048 360 bp mRNA linear EST 23-JAN-2001  
 DEFINITION IL5-GN0178-311000-204-f10 GN0178 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF993048  
 VERSION BF993048.1 GI:12399371  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 360)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

## sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## MEDLINE

20202663

## COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-GN0178-311000-204-f10&t3=2000-10-31&t4=1>)

Seq primer: puc 18 forward

High quality sequence stop: 305.

## FEATURES

## source

1. .360  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="GN0178"  
 /dev\_stage="Adult"  
 /note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI  
 ; Site\_2: SmaI; A mini-library was made by cloning  
 products derived from ORESTES PCR (U.S. Letters Patent  
 application No. 196,716 - Ludwig Institute for Cancer  
 Research) profiles into the pUC 18 vector. Reverse  
 transcription of tissue mRNA and cDNA amplification were  
 performed under low stringency conditions."  
 59 a 80 c 142 g 78 t

## BASE COUNT

## ORIGIN

## alignment\_scores:

Quality: 625.00 Length: 119  
 Ratio: 5.342 Gaps: 0

Percent Similarity: 98.319 Percent Identity: 98.319

## alignment\_block:

US-10-041-770-2 x BF993048/rev

```

Align seg 1/1 to reverse of: BF993048 from: 1 to: 360
158 GlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnArgGhi 174
|||||
358 GGTATGGAGAGTGCCCTNTGCAATTCACATGCGACCGGAACCGCAGGCA 309
|||||
174 SProArgSerProProArgSerGluLeuSerLeuLeuSerSerArgGlyG 191
|||||
308 CCTCGGAGCCACCACGATGAGTGTCCCTGATCTTCTAGAGGGG 259
|||||
191 LuGluAlaIleProSerProThrProArgAlaGluProPheSerAlaAsn 207
|||||
258 AAGAGCCTATTCGTCCTCCCTACTCCAAAGACGACAGCCATTCCTCCGAAAC 209
|||||
208 GlySerProGlnThrGluLeuProProThrGluLeuSerValHisThrPr 224
|||||
208 GCGAGCCCAACAACTAGTGTCTCCACAGAACTGTCTGTCCACACCCC 159
|||||
224 OSerProGlnAlaGluProLeuSerProGluThrAlaGlnThrGluValA 241
|||||
158 ATCCCCCAAGCAGACACCTCTAAGCCCTGAACTGCTCAGACAGAGGTGG 109
|||||
241 LaProArgThrArgProAlaProLeuArgHisHisProArgAlaGlnAla 257
|||||
108 CCCCCAGAACCGGCGTCCCTCCACAGAACTGTCTGTCCACACCCC 59
|||||
258 SerGlyThrGluProProSerProThrHisSerLeuGlyGluGlyPh 274
|||||
58 TCTGGCAGACAGCCCCCTCACCAGCAGCACTCCTTAGGAGAGGTGGCTT 9
|||||
274 ePheArg 276
|||||
8 CTTCCGT 2

seq_name: gb_est2:BE757541

seq_documentation_block:
LOCUS BE757541 571 bp mRNA linear EST 25-APR-2001
DEFINITION 212006 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE757541
VERSION BE757541.1 GI:10171533
KEYWORDS EST.
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.I., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteira,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGGTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 64 row: N column: 11
Seq primer: ATTGAGTGACACTATAG.
Location/Qualifiers

FEATURES

```

source

1. 571  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH108"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."  
96 a 177 c 209 g 89 t

BASE COUNT  
ORIGIN

alignment\_scores:  
Quality: 606.00 Length: 113  
Ratio: 5.611 Gaps: 0  
Percent Similarity: 95.575 Percent Identity: 92.920

alignment\_block:  
US-10-041-770-2 x BE757541 ..  
Align seg 1/1 to: BE757541 from: 1 to: 571

745 HisArgGlnLeuGlnCysArgGlnGluPheGlyGlyGlySerSerVa 761  
|||||  
3 CACCGTCAGCTACCTGCGCGCAGGAGTTTGGGGTGGCGGCTCCTCAGT 52  
|||||  
761 LProProGluArgCysGlyHisLeuProArgProAsnIleThrGlnSerC 778  
|||||  
53 GCCCTAGAGCGCTGCGGGCACCTGCCCGACCCACCAACATCACCCAGCCCT 102  
|||||  
778 ySglnLeuArgLeuCysGlyHisTrpGluValGlySerProTrpSerGln 794  
|||||  
103 GCCAGTCGCGCTCTGTGCCATTGGGAGGTTCGCTCACCTGGAGTCAG 152  
|||||  
795 CysSerValArgCysGlyArgGlyGlnArgSerArgGlnValArgCysVa 811  
|||||  
153 TGCTCTGTGCGATGCGGGCGGTCAGAGAGCGCGGAGTCCGCTGTGT 202  
|||||  
811 lGlyAsnAsnGlyAspGluValSerGluGlnGluCysAlaSerGlyProp 828  
|||||  
203 CGGCAACAACGGGATGAAGTGAGTGTGAGCAGGAGTGGCGCTCAGGGGCC 252  
|||||  
828 roGlnProProSerArgGluAlaCysAspMetGlyProCysThrThraLa 844  
|||||  
253 CGCGGCCCCCGCAGAGAGAGCGCTGCGACATGGGGCTGCACCCAGGCC 302  
|||||  
845 TrpPheHisSerAspTrpSerSerLysValSerProGlu 857  
|||||  
303 TGGTTCCACAGCGACTGGAGCTCCCAACTGCTCAGCGGAG 341  
|||||

seq\_name: gb\_est2:BG005397

seq\_documentation\_block:  
LOCUS BG005397 370 bp mRNA linear EST 24-JAN-2001  
DEFINITION PM1-GN0180-241100-003-e05 GN0180 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG005397  
VERSION BG005397.1 GI:12447519  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 370)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsumura,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMI&t2=PMI-GN0180-  
241100-003-e05&t3=2000-11-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 370.

FEATURES  
source  
1..370  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0180"  
/dev\_stage="Adult"  
/note="Organ: Placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT  
ORIGIN  
80 a 127 c 98 g 64 t

alignment\_scores:  
Quality: 594.00 Length: 117  
Ratio: 5.165 Gaps: 1  
Percent Similarity: 98.291 Percent Identity: 98.291

alignment\_block:  
US-10-041-770-2 x BG005397 ..

Align seg 1/1 to: BG005397 from: 1 to: 370

108 ProGluThrLeuProLeuTyr.ArgThrGlnSerArgGlyArgGlyGlyP 124  
|||||  
21 CCAGAAACCCCTCCCTNGTACAGGACACAGTCTCGGGAGGGGTGGCC 70  
|||||  
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluIle 140  
|||||  
71 CACTTCGAGTCCCGTCTCCACCTAGGAGAGAGAGAGAGAGAGAT 120  
|||||  
141 ArgAlaAlaArgSerArgLeuArgAspProIleLysProGlyMetPh 157  
|||||  
121 CGAGCGCCAGGAGGTCCCGGCTCGAGACCCCATCAAGCCAGGAATGT 170  
|||||  
157 eGlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnArgH 174  
|||||  
171 CGGTTATGGAGAGTGCCTTTTCATGTGCACACCGGACCGGAGGC 220  
|||||  
174 isProArgSerProProArgSerGluLeuSerLeuIleSerSerArgGly 190  
|||||  
221 ACCCTCGAGCCACCCAGATCTGAGCTGTCCCTGATCTCTTCTAGAGG 270  
|||||  
191 GluGluAlaIleProSerProThrProArgAlaGluProPheSerAlaAs 207  
|||||  
271 GAAGAGGCTATTCCGTCCTTACTCAAGAGCAGAGCCATCTTCCGCAAA 320  
|||||  
207 nGlySerProGlnThrGluLeuProProThrGluLeuSerValHisThr 223  
|||||  
321 CGGACGCCCCCAACCTGAGTCCCTCCACAGAACTGTGTGTCCACACC 369  
|||||

seq\_name: gb\_est2:BG900891

seq\_documentation\_block:  
LOCUS BG900891 449 bp mRNA linear EST 06-NOV-2001  
DEFINITION HOA7-1-H11 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION BG900891  
VERSION BG900891.1 GI:14311140  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 449)  
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
Lark,M.W.  
TITLE Identification and initial characterization of 5000 expressed  
sequenced tags (ESTs) each from adult human normal and  
osteoarthritic cartilage cDNA libraries  
Osteoarthritic Cartilage 9 (7), 641-653 (2001)  
JOURNAL 21482651  
MEDLINE  
COMMENT Contact: Sanjay Kumar  
UN2109  
Glaxosmithkline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@gsk.com  
Seq primer: T7  
FEATURES  
source  
1..449  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HOA (Human Osteoarthritic Cartilage)"  
/tissue\_type="cartilage"  
/lab\_host="E.coli DH10 B"  
/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
Directional"  
BASE COUNT 71 a 158 c 134 g 86 t  
ORIGIN

alignment\_scores:  
Quality: 579.00 Length: 106  
Ratio: 5.567 Gaps: 0  
Percent Similarity: 98.113 Percent Identity: 96.226

alignment\_block:  
US-10-041-770-2 x BG900891 ..

Align seg 1/1 to: BG900891 from: 1 to: 449

1 MetGluAsnTrpThrGlyArgProTrpLeuTyrLeuLeuLeuLeuLeu 17  
|||||  
132 ATGGAGAACTGGACTGGCAGGCCCTGGCTGTATCTGCTGCTCTGTC 181  
|||||  
17 rleuProGlnLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeu 34  
|||||  
182 CTTCCCTCAGCTCTCTTGGATCAGAGGTGTGTCCCCACACTCTCTTC 231  
|||||  
34 InThrProThrGluGlyGlnGlyProGluGlyValTrpGlyProTrp 50  
|||||  
232 AGACACCTACAGAGGAGGCCAGGCCCGAAGGTGTCTGGGACCTTGG 281  
|||||  
51 ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgAr 67  
|||||  
282 GTCCAGTGGGCTCTTGTCCAGCCCTCGGGGTGGGGTGCAGCGCAA 331  
|||||  
67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84  
|||||  
332 GAGCCGGACATGTCAGTCCCTCCTACAGTGCAGCTTCACCCGAGTCTGCC 381  
|||||  
84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100  
|||||

```
||||| |||||||||||||||||||||||||||||||||||||||
382 TCCTTCCGGGCCCCCAAGACATCCAGAGCCCTTCTTCCCGGGGCCAG 431
101 GlyProArgProGlnThr 106
|||||:|||||||||
432 GTCCCAACCCCGAGACT 449
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:40:30 ; Search time 292.16 Seconds  
(without alignments)  
15479.027 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atggagaactgactggcag.....octcagccttcacgcatag 2634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1259	47.8	2680	22 AAC91178	ADAM gene #2. Hom
2	472.4	17.9	1020	22 AAD16773	Human novel protei
3	472.4	17.9	1998	22 AAD16758	Human novel protei
4	165.2	6.3	2964	22 AAH16636	Human cdna sequenc
5	118.2	4.5	704	22 AAH99398	Human protein enco
6	113.2	4.3	2805	22 AAS97176	Human metalloprote
7	106.8	4.1	3675	24 AAS97174	Human metalloprote
8	98.2	3.7	5338	22 AAH49372	Human metalloprote
9	97	3.7	9295	23 ABL02495	Drosophila melanog

10	97	3.7	21314	23 ABL02494	Drosophila melanog
11	96.6	3.7	5061	22 AAF82149	Human ADAM type me
12	96.2	3.7	5353	22 AAD04475	Human 27875 ADAM-T
13	95.6	3.6	2274	22 AAD07967	Human ADAM type me
14	95.6	3.6	2727	22 AAD07956	Human ADAM type me
15	95.6	3.6	2879	22 AAH49370	Human metalloprote
16	95.6	3.6	3160	22 AAD07968	Human metalloprote
17	95	3.6	5353	22 AAD04494	Human ADAM type me
18	94	3.6	2848	22 AAF63439	Human ADAM type me
19	92	3.5	3776	24 AAH43769	ADAMTS-E DNA. Hom
20	91.6	3.5	3312	24 AAF82157	Human ADAM type me
21	91.6	3.5	3312	24 AAS97173	Human metalloprote
22	91.4	3.5	3199	22 AAS02432	Human cdna encodin
23	89.8	3.4	2575	22 AAD17501	Human ADAMTS-J1.4
24	89.4	3.4	814	22 AAH08271	Human cdna clone (
25	88.4	3.4	3409	22 AAF63448	Human ADAMTS-10 al
26	85.4	3.2	6505	21 AAF75588	Human OREFX ORF143
27	80.6	3.1	4287	24 AAS97177	Human metalloprote
28	77.4	2.9	1119	22 AAD14366	Novel human protei
29	77.4	2.9	1953	22 AAD14368	Novel human protei
30	77.4	2.9	2316	22 AAD14370	Novel human protei
31	77.4	2.9	4854	22 AAD14371	Novel human protei
32	74.2	2.8	4067	21 AAA95823	Human metalloprote
33	74	2.8	3218	22 AAF63440	Human ADAMTS-7 CDN
34	73.6	2.8	1698	22 AAD17500	Human ADAMTS-J1.3
35	73.6	2.8	1966	22 AAD17502	Human ADAMTS-J1.3
36	73.4	2.8	3263	22 AAH49371	Human ADAMTS-10 CD
37	73.4	2.8	3885	22 AAF63445	Human ADAMTS-10 CD
38	73	2.8	846	22 ABA08559	Human HSPp-124 pro
39	73	2.8	1341	22 AAD14365	Novel human protei
40	73	2.8	2175	22 AAD14367	Novel human protei
41	73	2.8	2538	22 AAD14369	Novel human protei
42	73	2.8	5076	22 AAD14364	Novel human protei
43	73	2.8	7260	22 AAD17757	Human novel KIRA12
44	73	2.8	8578	22 AAD14372	Novel human protei
45	72.6	2.8	1905	22 AAD17499	Human ADAMTS-J1.2

ALIGNMENTS

RESULT	ID	AC	XX	XX	DT	DE	XX	KW	KW	OS	PN	XX	PD	XX	PF	XX	PR	PR	XX	XX	PI	XX	DR	XX	PT
1	AAC91178																								
	AAC91178																								
	AAC91178																								
	20-MAR-2001																								
	ADAM gene #2.																								
	ADAM; disintegrin; metalloprotease; food additive; breast cancer;																								
	ovarian; ss.																								
	Homo sapiens.																								
	WO200073323-A2.																								
	07-DEC-2000.																								
	25-MAY-2000; 2000WO-US14308.																								
	27-MAY-1999; 99US-0136388.																								
	09-JUL-1999; 99US-0142930.																								
	28-JAN-2000; 2000US-0178717.																								
	{HUMA-} HUMAN GENOME SCI INC.																								
	Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;																								
	WPI; 2001-016507/02.																								
	Seven nucleic acid molecules encoding ADAM polypeptides containing a																								

A A

PT disintegrin and metalloprotease domain, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -

PS Claim 1; Page 266; 287pp; English.

CC The present invention relates to seven members of the ADAMs (proteins  
 CC which contain a Disintegrin And Metalloprotease domain) protein family.  
 CC The ADAMs proteins and DNA may be used to treat disease, as a food  
 CC additive or preservative, for chromosome identification, as probes  
 CC for diagnosing a disorder related to the female reproductive system,  
 CC particularly breast and/or ovary cancer. They are also useful in the  
 CC gene therapy of breast and ovarian cancer.

XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

Query Match 47.8%; Score 1259; DB 22; Length 2680;

Best Local Similarity 100.0%; Pred. No. 8.9e-287;

Matches 1259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 ctgaaagggtccaggatgggacccctgtgtcagccttgaggcccttgacatctgtgtggtg 1360  
 DB 3 ctgaaagggtccaggatgggacccctgtgtcagccttgaggcccttgacatctgtgtggtg 62  
 QY 1361 gacgtgtctgagcccggtgtgtggtggtctgtggtctgtgagcgtctcctgagggct 1420  
 DB 63 gacgtgtctgagcccggtgtgtggtggtctgtggtctgtgagcgtctcctgagggct 122  
 QY 1421 gtggagtcctgtgggggtgagttacattacattcgtcgcctgttttcgggaaacctcaactgaac 1480  
 DB 123 gtggagtcctgtgggggtgagttacattacattcgtcgcctgttttcgggaaacctcaactgaac 182  
 QY 1481 gaggggggccctcgtgggtacaaagatctgtggtacagcgggagccttcggtcc 1540  
 DB 183 gaggggggccctcgtgggtacaaagatctgtggtacagcgggagccttcggtcc 242  
 QY 1541 agattgccagtcctcgtcctagctccaaactacattcgtgcactctgtgcccctggggcggt 1600  
 DB 243 agattgccagtcctcgtcctagctccaaactacattcgtgcactctgtgcccctggggcggt 302  
 QY 1601 ccattcatcaatgggaactggcgtgtggtacccctcgtggtcctacagggcgcgagaccg 1660  
 DB 303 ccattcatcaatgggaactggcgtgtggtacccctcgtggtcctacagggcgcgagaccg 362  
 QY 1661 tctttcgatataacgctcctccaggaggagggagggagggagagtgctgtcgccggaag 1720  
 DB 363 tctttcgatataacgctcctccaggaggagggagggagggagagtgctgtcgccggaag 422  
 QY 1721 gccaccaccagcctgtggtgtgtatgtatgtattcagggagggaaacccaggggttt 1780  
 DB 423 gccaccaccagcctgtggtgtgtatgtatgtattcagggagggaaacccaggggttt 482  
 QY 1781 ttatcagtagtgcattctctcactcctccaaatccttgagaacccacccagagcccc 1840  
 DB 483 ttatcagtagtgcattctcactcctccaaatccttgagaacccacccagagcccc 542  
 QY 1841 ctgtcccccagctcagccgagattctgaggggtgagcccccaactgtctcggaccoc 1900  
 DB 543 ctgtcccccagctcagccgagattctgaggggtgagcccccaactgtctcggaccoc 602  
 QY 1901 gccagcccgagccagggcacccctcagcgtcaggtgaggtatccccagatgcccgcc 1960  
 DB 603 gccagcccgagccagggcacccctcagcgtcaggtgaggtatccccagatgcccgcc 662  
 QY 1961 cggccatccaggaacccctcgggtctcagctcgtactcgtactggaacagtgaggacat 2020  
 DB 663 cggccatccaggaacccctcgggtctcagctcgtactcgtactggaacagtgaggacat 722  
 QY 2021 ctgcatgctcagcgtctcgtcggaagaggtgtctggtcgcccccattttcctctgcatctcc 2080  
 DB 723 ctgcatgctcagcgtctcgtcggaagaggtgtctggtcgcccccattttcctctgcatctcc 782

QY 2081 gtgagtcgggagagaactggatgaacgagctgtgcgcgggtgccaggccccagcct 2140  
 DB 783 gtgagtcgggagagaactggatgaacgagctgtgcgcgggtgccaggccccagcct 842  
 QY 2141 cccctgaacccctgcacggcaccatgccccatactggagctggagctggcagtgacat 2200  
 DB 843 cccctgaacccctgcacggcaccatgccccatactggagctggagctggcagtgacat 902  
 QY 2201 cctgcagcgcctcctgtgtgcccccgccacccagcagcagcagctgcagtcgagtcgaggaat 2260  
 DB 903 cctgcagcgcctcctgtgtgcccccgccacccagcagcagcagctgcagtcgagtcgaggaat 962  
 QY 2261 ttggggggggtgctcctcctggtgcccccgagcgtgtggacatctccccggcccaaca 2320  
 DB 963 ttggggggggtgctcctcctggtgcccccgagcgtgtggacatctccccggcccaaca 1022  
 QY 2321 tcaccagctcttcagcagctgcctctgtggccattgggaagtgtgctctcttgagcc 2380  
 DB 1023 tcaccagctcttcagcagctgcctctgtggccattgggaagtgtgctctcttgagcc 1082  
 QY 2381 agtgcctcgtgcggtgcggccgggcccagagagcgcgaggttcgctgtgttggaaca 2440  
 DB 1083 agtgcctcgtgcggtgcggccgggcccagagagcgcgaggttcgctgtgttggaaca 1142  
 QY 2441 acgggtgatgaagtgcagcagcagaggtgtgcagcagccccccacagccccccagcag 2500  
 DB 1143 acgggtgatgaagtgcagcagcagaggtgtgcagcagccccccacagccccccagcag 1202  
 QY 2501 aggcctgtgacatggggccctgtactactgctgttcacagcagactggagctccaag 2559  
 DB 1203 aggcctgtgacatggggccctgtactactgctgttcacagcagactggagctccaag 1261

# RESULT 2

AD16773  
 ID AAD16773 standard; cDNA; 1020 BP.

XX AAD16773;

XX 22-NOV-2001 (first entry)

XX Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.

XX Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
 KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
 KW autoimmune disease; rheumatoid arthritis; pulmonary disorder; ischaemia;  
 KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
 KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
 KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
 KW anti-sense therapy; endocrine disorder; leukaemia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 65..847

XX /tag= a

XX /product= "Human novel protein"

XX /trans\_except= (pos:767..769, aa:Xaa)

XX /note= "CDS does not include start codon; Xaa is an

XX unknown amino acid"

XX /partial

XX WO200155202-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01325.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 27-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0215647.  
PR 07-JUL-2000; 2000US-0216680.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-451925/48.  
DR P-PSDB; AA09713.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT medical disorders and also for testing and detection e.g. diagnosis and  
PT screening for agonists -  
XX  
PS Claim 1; SEQ ID No 34; 469pp; English.  
XX  
CC AAD16750-AAD16775 represent cDNAs corresponding to novel human protein



PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-02556719.

PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-451925/48.  
XX P-PSDB; AAE09698.  
DR Isolated polypeptide for treating, preventing and/or prognosing  
DR medical disorders and also for testing and detection e.g. diagnosis and  
XX screening for agonists -  
PT Claim 1; SEQ ID No 19; 469pp; English.  
PS AAD16750-AAD16775 represent cDNAs corresponding to novel human protein  
XX genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777  
CC -AAD16780 represent novel human genomic DNA fragments. The novel proteins  
CC and their DNAs are useful for diagnosing, treating, preventing and/or  
CC prognosing inflammatory disorders (bursitis or tendonitis); neural  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system  
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);  
CC muscular disorders; reproductive disorders; gastrointestinal disorders  
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;  
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);  
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous  
CC disease and conditions (breast cancer); hyperproliferative disorders  
CC (leukaemia, hyperplasia); tumours; foetal and developmental  
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,  
CC asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine  
CC disorders; pregnancy-related disorders and infections. The novel protein  
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can  
CC also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
  
Query Match 17.9%; Score 472.4; DB 22; Length 1998;  
Best Local Similarity 99.6%; Pred. No. 1.6e-101;  
Matches 484; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
  
QY 2074 atctccgtgagtcgggagaggaactggtgaacgacgactgtgcgcgggtgcagagcc 2133  
Db |||||  
QY 2134 ccagctccctgaacctgccagggcaccatgcccccatactgggaggtggcgag 2193  
Db |||||  
QY 2194 tggacatctgcagcgcgtctctgtggcccgccacccagccagcagctgagtgccgg 2253  
Db |||||  
QY 2254 caggaatttgggggtggctctcctgcgtgccccgagcgtctggaacatctcccgg 2313  
Db |||||  
QY 2314 cccaacatcacccagctcttgcagctgcgcctctgtggccattgggaagtggctctct 2373  
Db |||||  
QY 2374 tggagccagtgctccgtgcgtgtggccggggccagagaaagccggcaggttcgtgtt 2433  
Db |||||  
QY 2434 gggaacaacggtgatgaagtgcagagcagagagtggtgcgtcagcccccacagccccc 2493  
Db |||||  
QY 2493 gggaacaacggtgatgaagtgcagagcagagagtggtgcgtcagcccccacagccccc 442

QY 2494 agcagagggcctgtgacatggggccctgtactactgctgctgttccacagcagctggagc 2553  
 Db |||||||  
 QY 443 agcagagggcctgtgacatggggccctgtactactgctgctgttccacagcagctggagc 502  
 Db |||||||  
 QY 2554 tccaag 2559  
 Db 503 tccaag 508

## RESULT 4

AAH16636  
 ID AAH16636 standard; cDNA; 2964 BP.

AC AAH16636;

DT 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:15752.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 8; SEQ ID 15752; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

XX represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

SQ Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

Query Match 6.3%; Score 165.2; DB 22; Length 2964;  
 Best Local Similarity 56.1%; Pred. NO. 4e-29;  
 Matches 333; Conservative 0; Mismatches 258; Indels 3; Gaps 1;

QY 1960 ccgcccatccagggacacccctgggtctccagctgctgactggaaacagagtgaggacac 2019  
 Db |||||  
 QY 2020 tctgatctcagctctctgggaaagggtgtctggcgccccattttctctctcatctcc 2079  
 Db |||||  
 QY 534 acagatgttccagacctgtgggaaaggatcgagctacccctatcttccgctgtgcac 593  
 Db |||||  
 QY 2080 cgtgagtcgggaggaactggatgaacgcagctgtccgctgggtgcccagcccccagcc 2139  
 Db |||||  
 QY 594 agaagcactcatgaagggtctctgaggttacttgactccagcatgaagcagaccccc 653  
 QY 2140 tccctgaacccctgccacggcacccccatgccccccatactggaggctggcgagtgagaca 2199  
 Db |||||  
 QY 654 gaggagggccctgcaacattctctctgcccagctcttgggacatcggggagtggtct 713  
 QY 2200 tctgagcgcctctgtgccccggccacccagcacccagctgagctgcggcaggaa 2259  
 Db |||||  
 QY 714 gagtgcagcaagacctgtgcccgggcatgcagcagcagcagctctgtgcccagggtg 773  
 QY 2260 ttgggggggtggtcctctcctggtgccccggagcgtgtggacatctccccggcccaac 2319  
 Db |||||  
 QY 774 tacgcaaccgcagctgacggtgacgctacccctgcccagcaccctggagaaacctgag 833  
 QY 2320 ataccagcttgcagctgcgctctgtggccattgggaagtgtgctctctctggagc 2379  
 Db |||||  
 QY 834 accaccagacctgccaactcaagatctgcagcagtggtggcagatccggaccactggacc 893  
 QY 2380 cagtcctccgtgcggtgcggcgccgagagagcggcgaggttcctgtgttgggaac 2439  
 Db |||||  
 QY 894 tcgtgctggtgcccctgcggcggtgggacagagaccctggtggaagtgtgagcaac 953  
 QY 2440 aacggtgatgaagtgcagcagcagagtgctgctcagggccccccacagccccccagaga 2499  
 Db |||||  
 QY 954 attgggatgtgttgacgatgaggaatgcaactgaagctc---cgccgaatgacatt 1010  
 QY 2500 gaggcctgtacatggggccctgactactgctgctggttccacagcagctggagc 2553  
 Db |||||  
 QY 1011 gagactgcagatgggacccctgtgcccagagcgtgttctccaccgagtgagc 1064

## RESULT 5

AAH99398

ID AAH99398 standard; cDNA; 704 BP.

AC AAH99398;

DT 16-OCT-2001 (first entry)

DE Human protein encoding cDNA sequence SEQ ID NO:233.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 antibacterial; endocrine; cardiac; central nervous system; virucide;  
 anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 antiaggregant; haemostatic; vulnery; antilicer; osteopathic; eczema;  
 dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
 genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 allergic rhinitis; diabetes; multiple sclerosis; depression;  
 Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 neurological disorder; ss.



XX WO200183782-A2.  
PN XX  
PD XX  
PF 08-NOV-2001.  
XX 04-MAY-2001; 2001WO-US14431.  
XX 04-MAY-2000; 2000US-201879P.  
PR XX  
XX (SUGEN-) SUGEN INC.  
PA XX  
XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
PI Payne V;  
PI  
XX WPI; 2002-041502/05.  
DR P-PSDB; AAU72891.  
DR  
XX Novel protease polypeptide useful for screening for substances that may  
PT be used to treat, e.g., cancers, immune-related diseases,  
PT cardiovascular disease, migraine, pain, psychotic and inflammatory  
PT disorders.  
XX  
PS Claim 30; Figure 1H; 232pp; English.  
XX  
CC The invention relates to an isolated, enriched, or purified protease  
CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to  
CC screen for substances (S) that may modulate its activity. Administering  
CC S (which modulates protease activity in vitro) may be used to treat a  
CC disease or disorder selected from cancers (e.g., of tissues, of blood or  
CC hematopoietic origin, of the breast, colon, lung, prostate, cervical,  
CC brain, ovarian, bladder or kidney), immune-related diseases and  
CC disorders, cardiovascular disease, brain or neuronal-associated diseases  
CC (e.g., central or peripheral nervous system diseases, migraine, pain,  
CC sexual dysfunction, mood disorders, attention disorders, neurological  
CC disorders, hypotension; hypertension, metabolic disorders, inflammatory  
CC disorders and dyskinesias), as a diagnostic tool for a disease or  
CC disorder such as those above. AAS97159-AAS97195 represent human  
CC protease coding sequences and primers of the invention.  
XX  
SQ Sequence 3675 BP; 874 A; 1009 C; 1019 G; 773 T; 0 other;

Query Match 4.3%; Score 113.2; DB 24; Length 2805;  
Best Local Similarity 50.0%; Pred. No. 6.9e-17;  
Matches 317; Conservative 0; Mismatches 308; Indels 9; Gaps 1;  
QY 1159 cgggcccgcagtcgagcccttaactccaggaattcattgcccagctgtatcagtg 1218  
DB 1039 cgggtcaacagtcgagataataacagcaaaccttccgtggatgtttaccagtg 1098  
QY 1219 gaggccttcaactgaagtcagcgtccagcagcgtgtaacgaactcgcggcccgctggc 1278  
DB 1099 aaacctatacaaaagtggaggaagatcgatcaaacactgactcaagcgtgagac 1158  
QY 1279 ttccgctttatgtccgtccactgaaaggctccaggatgggacccctgtgtaagcctgga 1338  
DB 1159 ttggaatttttttgcgaatgtccggcaagtgaaagtgaaactccctgtccccaac 1218  
QY 1339 gccctgacatctgtggtgagcgtgtgagcccggtgtgagtggtggtcctggc 1398  
DB 1219 aaaaatgatgtttgatgacgggtttgtgaactagtggaatggtgatacgaactagc 1278  
QY 1399 tctgacagcgtcctgatggctgtgagtcgtgtggtgggtgagtgatctacactgtgcctt 1458  
DB 1279 tctaaagcagtttccagatctgtgtggtttgcaaaagtgataatcaactgtcaagttt 1338  
QY 1459 gtttcggggaacctcactgacagcagggggccctcgtggctatcagaagatctgtggatt 1518  
DB 1339 tataaagcctgtactcaaccagcataaagcaaatgaattatccggtgtcctcatt 1398  
QY 1519 ccagcggagccttgcgtccagattgccagctcccgctcagtcagtcacactgagcga 1578  
DB 1399 ccagctggccgcccagcatcgaaatccaggagctgcaggtttctccagttacatcgca 1458  
QY 1579 cttcgtgcccctggggcggtccatcattcaatggaactgggtgtggtatccctggg 1638  
DB 1459 gttcgaagcctcagtcgaagattaccctaccggggcgtggagcagctgactggcctggg 1518  
QY 1639 tectacagggcggcgggaccgtctttgatataaccgtctcccgaggagggagggcaaa 1698  
DB 1519 gattcccccttcgctgggaccacgtttgaataccagcgtctttcaa-----ccgc 1569  
QY 1699 ggggagagctctgctgggtgaagcccccacacccagcgtgtggtatgtatgtatttt 1758  
DB 1570 ccggagcgtctgacgcgcagggcccaaaatgagacgctggtttgaaattctgatg 1629  
QY 1759 caggaggaacccagcgttttttatcagtag 1792  
DB 1630 caaggcaaaatccaggagtagcttggaagtatg 1663  
RESULT 7  
AAS97174  
ID AAS97174 standard; cDNA; 3675 BP.  
XX  
AC AAS97174;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Human metalloprotease partial DNA sequence #3.  
XX  
KW Human; protease; PCR primer; cytosolic; immunomodulator; cardiac;  
KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;  
KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;  
KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;  
KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;  
KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;  
KW immune-related disease; cardiovascular disease; neuronal disease;  
KW migraine; sexual dysfunction; mood disorder; attention disorder;  
KW cognition disorder; hypotension; hypertension; psychotic disorder;  
KW dyskinesia; metabolic disorder; inflammatory disorder; ss.  
XX Homo sapiens.



Db 2262 gcaccacacacaccagattatcatcatgtcaccattctcttggagcccgagtat 2321  
Qy 1539 ccagattgccagctccgcttagctccaaactacctggcaactctgtggccctggggccg 1598  
Db 2322 ccgcatcatgaatgaacgtctctacctctcatcttctgtgcgaatgccctcagaag 2381  
Qy 1599 gtccatcatcaatgggaactggctgtggatccccctgggtccctacagggccggcgag 1658  
Db 2382 gtactacctgaatggcactgacccgtgactggcccgccggtacaaattttcgggcac 2441  
Qy 1659 cgtctttogataaaccttccctccagggagggagggcgaagggtgctgtcggtga 1718  
Db 2442 tactttogactacacaggtctctataagag-----cccgagaacttaactgctac 2492  
Qy 1719 agggcccccaccccgctgtgtgtgtcttatgtatcttccagggagggaaaccccggt 1778  
Db 2493 tggacacacacagacactgattgtggagctgtgttccaggaaggaaacccgggtg 2552  
Qy 1779 tttttatcagtatgtcatct 1800  
Db 2553 tgcctgggaataactcatgcct 2574

RESULT 8  
AAH49372  
ID AAH49372 standard; DNA; 5338 BP.  
XX  
AC AAH49372;  
XX  
XX 07-DEC-2001 (first entry)  
DE Human metalloprotease MPTS-19 encoding DNA.  
KW Metalloprotease; MPTS-10; human; thrombospondin domain; spondylitis;  
KW aggrecan associated disease rheumatoid arthritis; osteoarthritis;  
KW osteopathic; antiarthritic; antipsoriatic; antirheumatic; psoriasis;  
KW sport injury; joint trauma; fibrosis; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..5073  
FT /\*tag= a  
FT /product= "MPTS-19"  
XX  
DE10107360-A1.  
XX  
XX 06-SEP-2001.  
XX 16-FEB-2001; 2001DE-1007360.  
XX 18-FEB-2000; 2000US-184152P.  
XX (HOPF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
XX Heller RA, Klonowski P, Zuo F;  
XX  
XX WPI; 2001-607955/70.  
XX P-PSDB; AAB86949.  
XX  
XX Novel metalloprotease containing a thrombospondin domain (MPTS protein)  
XX is useful to treat aggrecan associated disease including rheumatoid  
XX arthritis and osteoarthritis .  
XX  
XX Claim 1b; Page 39-43; 56pp; German.  
XX  
XX This invention describes a novel metalloprotease containing a  
XX thrombospondin domain (MPTS protein) which is useful to treat aggrecan  
XX associated disease including rheumatoid arthritis and osteoarthritis.  
XX The products of the invention have osteopathic, antiarthritic,  
XX antipsoriatic and antirheumatic activity. MPTS may be used to identify a  
XX modulator of its activity, e.g., an agonist or antagonists. Such  
XX compounds, as well as MPTS itself may be used to treat disease

CC associated with MPRS activity or aggrecan degradation, particularly  
CC osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport  
CC injury, joint trauma or fibrosis. This sequence encodes the human  
CC MPTS-19 metalloprotease described in the method of the invention.  
XX  
SQ Sequence 5338 BP; 974 A; 1825 C; 1638 G; 901 T; 0 other;  
  
Query Match 3.7%; Score 98.2; DB 22; Length 5338;  
Best Local Similarity 51.5%; Pred. No. 2.8e-13;  
Matches 369; Conservative 0; Mismatches 318; Indels 30; Gaps 5;  
  
Qy 1090 tgttctggggagagtgaacagctgaagcctgcagcaagcgcctgcctcccccgtgacag 1149  
Db 1741 tgttgggtggagcagcgtctccctgcctgcacactgcagcctgcctgcctgcctgcctgc 1800  
Qy 1150 ccagaccctggccctgcagtcgagccttttaactcccaggaattcatgggccaactg 1209  
Db 1801 cctctctccgcacgtccagtcgacactttgacccatgctctacaaagggccgctg 1860  
Qy 1210 taccagtgaggagcccttcaactgaagtcagggtcccagcgtctgtaactgaactgcagg 1269  
Db 1861 cacacatgggtgcccgtg-----gtcaatgaactgaacccctgcgagctgcactgcgg 1914  
Qy 1270 cccgtggtctccgtctctatgtcgtcacactgaaaggtcccaggaatgggacccctgtg 1329  
Db 1915 ccgcgaatgagtactttgcgagaagctgcggagccctggtctgagtgcacccctgc 1974  
Qy 1330 cagcctggagc-----ccctgacatctgtgtggctggagcgtgtctgagcccccgtgt 1383  
Db 1975 taccaggtccgcagccgagcctctgtatcaacgcgcatctgtaagaactgtggctgt 2034  
Qy 1384 gatgggacctctgcttgcagcgcctcctgatgctgtgagctgtgtgggtgtgatgat 1443  
Db 2035 gacttcgagattgactcgcgtgtatgagagccgctgtgtgtgtgtccacggcgaacggc 2094  
Qy 1444 tctactgtcctgttctgttgggggaacctactgaccgagggggcccccctgggtatcag 1503  
Db 2095 tccacctgccacacgtgagcgggacccctgcagagagggccgaggg---cctgggggtatgtg 2151  
Qy 1504 aagatctgtgattccagcgggagccttgcgctccagattcccagctcccgccctcagc 1563  
Db 2152 gatgtggggctgatccacggcggcgagcagcagagatccgcataccaaagggtgtccgagct 2211  
Qy 1564 tccaaactactggcaactctgtggccctg---ggggccgggtccatcatcaatgggaactgg 1620  
Db 2212 gccaaactctctggcac:gcggagcgaggaccgcggaggaagtaactctcaatgtgtggctg 2271  
Qy 1621 gctgtggatccccctgggtcctacagggcggcgagcgtcttcttcgataaaccgtcct 1680  
Db 2272 accatccagtggaaacggggactaccagtggtggcagggaccaccccttcacatcacgcagcag- 2330  
Qy 1681 cccagggaggggagggagagagctgtcggctgaaggcccccaccacccagcctgtg 1740  
Db 2331 -----gggcaactgggagaacctcaactcctcccggttccacaaagggagcctgtc 2379  
Qy 1741 gatgtctatgatcttccagggaggaacccagggcgttttttcatcagtatgtcatc 1797  
Db 2380 tggatccagctgtgttccaggagagcaacctgggtgcaactacgagtagtaccatc 2436

RESULT 9  
ABL02495  
ID ABL02495 standard; cDNA; 9295 BP.  
XX  
AC ABL02495;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 1967.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.







Db 2403 tggatccagctgctgttccagagagcaacctgggtgcactacagatcacaccatc 2459

RESULT 13

AAAD07967

ID AAD07967 standard; cDNA; 2274 BP.

XX AC

XX AAD07967;

XX AC

XX 04-AUG-2001 (first entry)

XX DE Human protease-related protein #12 cDNA.

XX KW Human; novel human protein; NHP; protease-related protein;

XX KW metalloproteinase; ADAM-TS6; therapy; drug screening;

XX KW symptomatic; phenotypic manifestation; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..2274

XX FT /\*tag= a

XX FT /product= "Human protease-related protein"

XX PN WO200142468-A1.

XX PD 14-JUN-2001.

XX XX 08-DEC-2000; 2000WO-US33362.

XX PF 09-DEC-1999; 99US-0169769.

XX XX (LEXI-) LEXICON GENETICS INC.

XX PI Donoho G, Turner CA, Friedrich G, Scoville J, Zambrowicz B;

XX PI Sands AT;

XX WPI: 2001-381688/40.

XX P-PSDB; AA053583.

XX DR Novel isolated human protease polynucleotide that shares structural

XX PT similarity with animal proteases such as ADAM-TS6, a zinc

XX PT metalloproteinase, useful in therapeutic, diagnostic and pharmacogenic

XX PT applications

XX PS Disclosure; Page 43-44; 48pp; English.

XX CC The present cDNA sequence encodes a novel human protein (NHP),

XX CC protease-related protein. The protease-related protein shares structural

XX CC similarity with animal proteases and particularly with metalloproteinases

XX CC such as ADAM-TS6. The protease-related protein is used in therapeutic,

XX CC diagnostic and pharmacogenic applications. It is also used for the

XX CC detection of mutant protease-related protein, or inappropriately

XX CC expressed protease-related proteins for the diagnosis of a disease. It is

XX CC also used for drug screening which is effective in the treatment of

XX CC symptomatic or phenotypic manifestations of perturbing the normal

XX CC function of protease-related protein in the body. It is also used in

XX CC conjunction with polymerase chain reaction to screen libraries, isolate

XX CC clones, and prepare cloning and sequencing templates. It is also used as

XX CC hybridization probe for screening libraries, and assessing gene

XX CC expression patterns.

XX XX Sequence 2274 BP; 648 A; 460 C; 559 G; 607 T; 0 other;

XX SQ

Query Match 3.6%; Score 95.6; DB 22; Length 2274;

Best Local Similarity 48.8%; Pred. No. 9.1e-13;

Matches 288; Conservative 0; Mismatches 299; Indels 3; Gaps 1;

QY 1090 tggctgggagagagtgaaagctgaagagctgcagcagcagcgcctggccctgagcag 1149

DB 1336 tgcctgggaaaggaacggtatcgtctgttaacacagatccatgccttgggttcc 1395

QY 1150 ccagacccccggccctgcagtcgagcccttaactccacgaattcattgggcccagctg 1209

DB 1396 cgagatttcgagagaacacagtgacagcttggaataatgccttccgaggaagat 1455

QY 1210 tatcagtgaggagcccttcaactgaagtcacaggtccacgagctgtggaactgacg 1269

DB 1456 tataactggaaacccctatactgaggtgggttaa---aacctgtgcattaaactgtg 1512

QY 1270 ccccggtgcctccgcttctatgtccgtcaactgaaagtcacagagatgggacccctgtg 1329

DB 1513 gctgaaggttataattctacactgaacgtgctcctgcgtgatcgatgggacccagtc 1572

QY 1330 cagcctggagccctgacatctgtgtgctggagcgtgtctgagcccgctgtgatggg 1389

DB 1573 aatcggtatcacggtatctgcacatcaatggagaatgcaagcacgtaggctgtgataat 1632

QY 1390 atccttgctctgcagggcgtccgtggtgtggagctctgtggtgggtgtgatctctacc 1449

DB 1633 atttgggtatcgtatgctagggaagatagatgctgagctgtggtgggacggaagcaca 1692

QY 1450 tgcgcttgttcggggaacccctcaactgacgagggggcccccctggcgtacagaagatc 1509

DB 1693 tgtgatccattgaaagggttcttcaatgattcactgcccaggggaggtacatgggaagt 1752

QY 1510 ttgtgattccagcgggagccttgcgtccagattgccagctcccgccctagctccaac 1569

DB 1753 gtgcagataccaagagctctgttcacatgaagttagaagaagtgcacatgcaagaac 1812

QY 1570 tacctggcacttcgtgcccctggggcggtccatcaatcaatgggaaactgggtgtgat 1629

DB 1813 tatattgctttaaactgaaaggagatgattactatattatgtgctgctgacctgtgac 1872

QY 1630 cccctgggtcctacagggcgcggtgacgtcttcttgatataaacctgcc 1679

DB 1873 tggcctaggaaattgatgttgctgggacagcttttcattacaagagacc 1922

RESULT 14

AAAD07956

ID AAD07956 standard; cDNA; 2727 BP.

XX AC

XX AAD07956;

XX DT 04-AUG-2001 (first entry)

XX DE Human protease-related protein #1 cDNA.

XX KW Human; novel human protein; NHP; protease-related protein;

XX KW metalloproteinase; ADAM-TS6; therapy; drug screening;

XX KW symptomatic; phenotypic manifestation; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..2727

XX FT /\*tag= a

XX FT /product= "Human protease-related protein"

XX PN WO200142468-A1.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33362.

XX PR 09-DEC-1999; 99US-0169769.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Donoho G, Turner CA, Friedrich G, Scoville J, Zambrowicz B;

XX PI Sands AT;

XX WPI: 2001-381688/40.

XX DR P-PSDB; AAD07956.

XX Novel isolated human protease polynucleotide that shares structural  
PT similarity with animal proteases such as ADAM-TS6, a zinc  
PT metalloproteinase, useful in therapeutic, diagnostic and pharmacogenic  
PT applications  
XX  
PS Claim 1; Page 28-29; 48pp; English.  
XX  
XX The present cDNA sequence encodes a novel human protein (NHP),  
CC protease-related protein. The protease-related protein shares structural  
CC similarity with animal proteases and particularly with metalloproteinases  
CC such as ADAM-TS6. The protease-related protein is used in therapeutic,  
CC diagnostic and pharmacogenic applications. It is also used for the  
CC detection of mutant protease-related protein, or inappropriately  
CC expressed protease-related proteins for the diagnosis of a disease. It is  
CC also used for drug screening which is effective in the treatment of  
CC symptomatic or phenotypic manifestations of perturbing the normal  
CC function of protease-related protein in the body. It is also used in  
CC conjunction with polymerase chain reaction to screen libraries, isolate  
CC clones, and prepare cloning and sequencing templates. It is also used as  
CC hybridization probe for screening libraries, and assessing gene  
CC expression patterns.  
XX  
SQ Sequence 2727 BP; 788 A; 546 C; 650 G; 743 T; 0 other;

Query Match 3.6%; Score 95.6; DB 22; Length 2727;  
Best Local Similarity 48.8%; Pred. No. 9.6e-13;  
Matches 288; Conservative 0; Mismatches 299; Indels 3; Gaps 1;  
QY 1090 tgtttctgggagagtgaaacagctgaagcctgacgacccgagcgcctccccctgagcag 1149  
DB 1789 tgccttggggaagaaacggtatgcctctgtaacacagatccatgcccttgggttcc 1848  
QY 1150 ccagacccccggccctgagtcgacgctttaaactcccagggaattcatggccagctg 1209  
DB 1849 cgagatttcgagaaacagtgacacttgacaatatgcttccgagaaagtat 1908  
QY 1210 tatcagttggagcccttcaactgaagtcacaggtccacagcctgtgaactgaactgcgg 1269  
DB 1909 tataactggaaacccatactgaggtgggttaa---aaccttggcattaaactgcttg 1965  
QY 1270 cccctggcttcgctctatgctccgtcacactgaaaggtccaggtgggaccctgtgt 1329  
DB 1966 gctgaaggtataattctacactgaacgtgctcctggtgatgagtgagccagctgc 2025  
QY 1330 cagctggagcccttgacatctgtgtggtgagcgtgtctgagcccgctgtgatggg 1389  
DB 2026 aatcggtattcactggtatgcatcaatggagaatgcaacgacgtaggtgtgataat 2085  
QY 1390 atccttgctctggcagcgctcctgatgctgtgagctgtggggtgatcttacc 1449  
DB 2086 atttgggatactgactaggaagaatagatgtcgtgtgtggaggggacgaagcaca 2145  
QY 1450 tgcgccttgttctggggaaacctcactgaccagggggcccccctggctatcagaagatc 1509  
DB 2146 tgtgatgcattgaaggggttctcaatgattcactgccaggggaggtcactatggaagt 2205  
QY 1510 ttgtgattccagcgggaacccctggcctcagattgccagctccgctagctccac 1569  
DB 2206 gtgcagatacccaagggtctgttcaactggaagttagaggttgccatgtccaagaac 2265  
QY 1570 taactggcacttgcctggcctggggccggtccatcataatgggaactgggctgtggat 1629  
DB 2266 tataatgctttaaaactgaagagatgattactataatgaaggtgcccggactattgac 2325  
QY 1630 cccctcgggtccctacagggccgggacgcttcttcgtataaacgtcc 1679  
DB 2326 tggccctaggaaattgatgtgtgctggacagcttttctataaagagacc 2375

RESULT 15  
AAH49370

ID AAH49370 standard; DNA; 2879 BP.  
XX  
AC AAH49370;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human metalloprotease MPTS-15 encoding DNA.  
XX  
KW Metalloprotease; MPTS-15; human; thrombospondin domain; spondylitis;  
KW aggrecan associated disease rheumatoid arthritis; osteoarthritis;  
KW osteopathic; antiarthritic; antipsoriatic; antirheumatic; psoriasis;  
KW sport injury; joint trauma; fibrosis; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2879  
FT /tag= a  
FT /product= "MPTS-15"  
XX  
XX DE10107360-Al.  
XX  
XX 06-SEP-2001.  
XX  
XX 16-FEB-2001; 2001DE-1007360.  
XX  
PR 18-FEB-2000; 2000US-184152P.  
XX  
XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
XX Heller RA, Klonowski P, Zuo F;  
XX  
XX WPI; 2001-607955/70.  
XX  
XX P-PSDB; AAB86947.  
XX  
XX Novel metalloprotease containing a thrombospondin domain (MPTS protein)  
XX is useful to treat aggrecan associated disease including rheumatoid  
XX arthritis and osteoarthritis  
XX  
XX Claim 1b; Page 24-27; 66pp; German.  
XX  
XX This invention describes a novel metalloprotease containing a  
XX thrombospondin domain (MPTS protein) which is useful to treat aggrecan  
XX associated disease including rheumatoid arthritis and osteoarthritis.  
XX The products of the invention have osteopathic, antiarthritic,  
XX antipsoriatic and antirheumatic activity. MPTS may be used to identify a  
XX modulator of its activity, e.g., an agonist or antagonists. Such  
XX compounds, as well as MPTS itself may be used to treat disease  
XX associated with MPTS activity or aggrecan degradation, particularly  
XX osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport  
XX injury, joint trauma or fibrosis. This sequence encodes the human  
XX MPTS-15 metalloprotease described in the method of the invention.  
XX  
SQ Sequence 2879 BP; 821 A; 574 C; 679 G; 805 T; 0 other;

Query Match 3.6%; Score 95.6; DB 22; Length 2879;  
Best Local Similarity 48.8%; Pred. No. 9.7e-13;  
Matches 288; Conservative 0; Mismatches 299; Indels 3; Gaps 1;

QY 1090 tgtctggggagtgaaacagctgaagcctgacgacccgctccccctgagcag 1149  
DB 1789 tgccttggggaagaaacggtatgcctctgtaacacagatccatgcccttgggttcc 1848  
QY 1150 ccagacccccggccctgagtcgacgctttaaactcccagggaattcatggccagctg 1209  
DB 1849 cgagatttcgagaaacagtgacacttgacaatatgcttccgaggaagtat 1908  
QY 1210 tatcagttggagcccttcaactgaagtcacaggtccacagcctgtgaactgaactgcgg 1269  
DB 1909 tataactggaaacccatactgaggtgggttaa---aaccttggcattaaactgcttg 1965  
QY 1270 cccctggcttcgctctatgctccgtcacactgaaaggtccaggtgggaccctgtgt 1329

Db 1966 gctgaaggttataattctacactgaactgctctcggtgatcgatggacccagtgc 2025  
QY 1330 cagcctggagccctgacatctgtgtggtgagcgtgtctgagccccggctgtgtgagg 1389  
Db 2026 aatgcggattcaactgatactgcatcaatgagaatgcaagcaactaggtgtgataat 2085  
QY 1390 atccttggctctggcagcgctctgatggctgtggagtgtgtgggggtgatattctacc 1449  
Db 2086 attttggatctgatctagggagatagatgtcgagtgtgtggagggacggaagcaca 2145  
QY 1450 tgtcgcttgttttcggggaaacctgaactgacccggggggccccctgggctatcagaagatc 1509  
Db 2146 tgtgatgccattgaagggttcttcaatgattcactgccaggggaggtacatgggaagtg 2205  
QY 1510 ttgtgaattccagcgagagccttgcggctccagattgccagctccggcctagctccaac 1569  
Db 2206 gtgcagataccaaagaggtctgttcacattgaagttagagaagttgccatgtcaagaac 2265  
QY 1570 taactggcaacttcgtggccctggggccgggtccatcataatgggaactgggctgtggat 1629  
Db 2266 tatattgctttaaaatctgaaggagatgattactatattaatgtgtgcctggactattgac 2325  
QY 1630 cccctgggtccctacagggccgggacccgtcttttcgatatataccgtcc 1679  
Db 2326 tggcctaggaaatttgatgtgtgtgggacagcttttcattacaagagacc 2375

Search completed: July 23, 2002, 17:26:49  
Job time: 6379 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:36:00 ; Search time 3257.62 Seconds  
(without alignments)  
16920.481 Million cell updates/sec

Title: US-10-041-770-1

Perfect score: 2634

Sequence: 1 atggagaactgactggcag.....cctcagcctttccagcatag 2634

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htgc\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No. Score Match Length DB ID Description

length of match

1	2508	95.2	4230	6	AX342636	AX342636 Sequence
2	699	26.5	176550	9	AL356356	AL356356 Human DNA
3	529	20.1	2070	9	AF217974	AF217974 Homo sapi
4	457	17.4	207815	2	AC053497	AC053497 Homo sapi
5	205	7.8	207815	2	AC053497	AC053497 Homo sapi
6	39	1.5	142023	2	AC092479	AC092479 Mus muscu
7	39	1.5	215810	2	AC093317	AC093317 Mus muscu
8	30	1.1	1409	10	BC016215	BC016215 Mus muscu
9	23	0.9	2703	9	BC008840	BC008840 Homo sapi
10	23	0.9	2964	9	AK023772	AK023772 Homo sapi
11	23	0.9	62385	2	AC101188	AC101188 Mus muscu
12	23	0.9	151333	2	AC026636	AC026636 Homo sapi
13	23	0.9	184483	2	AC104938	AC104938 Homo sapi
14	22	0.8	168872	2	AC019036	AC019036 Homo sapi
15	22	0.8	175970	9	AC097488	AC097488 Homo sapi
16	21	0.8	4506	2	AC103667	AC103667 Homo sapi
17	21	0.8	45257	2	AC084177	AC084177 Homo sapi
18	21	0.8	67047	2	AC013544	AC013544 Homo sapi
19	21	0.8	72907	2	AC016010	AC016010 Homo sapi
20	21	0.8	138706	9	AP003469	AP003469 Homo sapi
21	21	0.8	148427	2	AC018981	AC018981 Homo sapi
22	21	0.8	168811	2	AC018941	AC018941 Homo sapi
23	21	0.8	168601	2	AC016803	AC016803 Homo sapi
24	21	0.8	183979	2	AL354671	AL354671 Homo sapi
25	21	0.8	189046	2	AL592301	AL592301 Homo sapi
26	21	0.8	209844	9	AC011495	AC011495 Homo sapi
27	21	0.8	215441	9	AC005736	AC005736 Homo sapi
28	21	0.8	1189	6	AX061802	AX061802 Sequence
29	20	0.8	1326	5	DR043898	DR043898 Danio rerio
30	20	0.8	1575	9	AK026179	AK026179 Mus muscu
31	20	0.8	1927	10	AB040490	AB040490 Mus muscu
32	20	0.8	1979	5	DR077736	DR077736 Danio rerio
33	20	0.8	2189	5	AF039952	AF039952 Danio rer
34	20	0.8	2239	9	AK057145	AK057145 Homo sapi
35	20	0.8	2562	9	BC018923	BC018923 Homo sapi
36	20	0.8	2671	9	AB056152	AB056152 Homo sapi
37	20	0.8	2994	5	DR2P12POU	DR2P12POU
38	20	0.8	3160	8	BEP8GEN	BEP8GEN
39	20	0.8	3215	10	RNU96920	RNU96920
40	20	0.8	3246	9	AF218313	AF218313 Homo sapi
41	20	0.8	3840	10	MMFBLN2S01	MMFBLN2S01
42	20	0.8	4398	10	RNU96921	RNU96921
43	20	0.8	38399	2	AL354885	AL354885 Homo sapi
44	20	0.8	47338	9	AL355990	AL355990 Human DNA
45	20	0.8	47338	9	AL355990	AL355990 Human DNA

#### ALIGNMENTS

RESULT 1

AX342636

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

FEATURES

Source

AX342636 Sequence 33 from Patent WO0198468. 4230 bp DNA linear PAT 12-JAN-2002

AX342636 human.

AX342636 Homo sapiens

AX342636 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AX342636 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AX342636 1 (sites)

AX342636 Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,

AX342636 Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,

AX342636 Hafalla, A., Khan, F.A., Walla, N.K., Yao, M.G., Lu, D.A., Patterson, C.,

AX342636 Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y., and Reddy, R.

AX342636 Patent: WO 01/98468-A 33 27-DEC-2001;

AX342636 Incyte Genomics, Inc. (US)

AX342636 Location/Qualifiers

AX342636 1..4230

AX342636 /organism="Homo sapiens"

AX342636 /db\_xref="taxon:9606"

AX342636 /note="Incyte ID No: 7604035CBI"



	BASE COUNT	810	a	1411	c	1210	g	799	t	
	ORIGIN									
	Query Match		95.2%	Score	2508;	DB	6;	Length	4230;	
	Best Local Similarity		100.0%;	Pred.	No.	0;				
	Matches	2558;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	
QY	1	atggagaactggactgcaggccctggctgttatctgctgtgcttcttgtctcctcctcaag	60							
Db	228	ATGGAGAANTGGACTGCGCAGGCCCTGGCTGTATCTGCTGCTGCTTCTTCCTCCCTCCCTCAG	287							
QY	61	ctctgcttgattcaggagtgtgttcggacacctctctcagacaacctcacagagagggc	120							
Db	288	CTCTGCTTGATCATCAGAGGTGTGTCGGACACTCTCTTCAGACACCTPACAGAGAGGGGC	347							
QY	121	caggcccccaaggtgtctcgggaaccttgggtccagtgggcotcttgtcccagccctgc	180							
Db	348	CAGGGCCCCGAAGGTGCTCTGGGACCTTTGGGTCCAGTGGGCTCTTGCTCCAGCCCTGC	407							
QY	181	ggggtgggggtcagcgacgagccgacatgtcagctccctacagtccagtcagctcacccg	240							
Db	408	GGGGTGGGGGTGCAGCCGACAGACCGGACATGTCAAGTCCCCTPACAGTCCAGCTCCACCCG	467							
QY	241	agttgccctccctcccgcggccccaagacatccagaagccctctcccgcggggccag	300							
Db	468	AGTCTGCCCTCCCTCCCGCGCCCCAAGACATCCAGAAAGCCCTCCTCCCGCGGGCCAG	527							
QY	301	ggtccagacccccagactctccagaaacctccocttgtacaggacacagtctcgggga	360							
Db	528	GGTCCAGACCOCAGACTTCTCCAGAAACCTTCCCTTTGACGGACACAGTCTCGGGGA	587							
QY	361	aggggtggccacttcaggtccggtctccacctcaggtagggagagagaccagagatt	420							
Db	588	AGGGGTGGCCCACTTCGAGSTCCCGCTTCCCACCTTAGGGAGAGGAGACCCAGAGATT	647							
QY	421	cgaagcggcaggaggtccggcttcgagaccocatcaagccaggaaatgttcggttatggg	480							
Db	648	CGAGCGGGCAGAGGFTCCCGGCTTCGAGACCCCATTAAGCCAGGAATGTTCCGTTATGGG	707							
QY	481	agagtgcctttgatctgcactgcaccggaaccgcagccctcgagagccaccacaga	540							
Db	708	AGAGTGCCTTTGCATTGCCATGTCAACGGGAACCGCAGCACCTTCGGAGGCCACCCAGA	767							
QY	541	tctgagctgcctgatctctctctagggggaagggtatttcogtccctactccaaga	600							
Db	768	TCTGAGCTCTCCCTGATCTCTTCTAGGGGAAGAGCCTATTCCGCTCCCTTACTCCAAGA	827							
QY	601	gcagagccattctccgaaaaggcagcccaaacctagctccctccacagaaactctc	660							
Db	828	GCAGAGCCATTCTCGCAAACGGCAGCCCCAACCTGAGCTCCCTTCCACAGAACTGTCT	887							
QY	661	gtccacaccccatcccccaagcagaacctataagccctgaaactgctcagacagaggtg	720							
Db	888	GTCCACACCCCATCCCCCAAGCAGAACTCTTAAGCCCTGAAACTGCTTCAGACAGAGGTG	947							
QY	721	gccccagaacaggacctgcaccttaaggcataccccagagccaggcocttggcaca	780							
Db	948	GCCCCAGAAACAGGCTGTCCCCCTTACGGCATCACCCAGAGCCGAGGCTCTGGCACA	1007							
QY	781	gagccccctcaccacgcactctcttaggaaggtggtcttcttcgtgatccccctcag	840							
Db	1008	GAGCCCCCTCACCCAGGCATCTCTTAGAGAAAGGTGCTTCTTCCGTGCATCCCTCAG	1067							
QY	841	ccagaaagccaaagtcccaagggttggccagtcaccaaggttagcagggagacgcctcgt	900							
Db	1068	CCACGAAGGCCAAGTTCACAGGTTGGGCCAGTCCCAAGTTAGCAGGGGACACGCCCTGTAT	1127							
QY	901	ccctttcttcggtccctcgggcccagagccagcagggccaaaggcccttgggaaacggg	960							
Db	1128	CCTTTTCTTCGTTCCGTCCTCGGGCCGAGGCCACAGAGGCCCAAGGGCCTTGGGGAACGGGG	1187							
QY	961	gggaactctcaaggcccccgcctggagcctgacctcagcaccggcgctggtgctgcc	1020							



```

QY 1033 gccccatgccagctccctctgagcctcttctgctccagtagccctattcccaagtgt 1092
Db 81286 GCCCCCATGCCAGCTCCCTCTGAGAGCTCTTGTCTCCAGTAGCCCTATTCAGATGT 81345

QY 1093 tctggagagagtgaaacagctaaagcctgcagccaagcg 1131
Db 81346 TCTGGGAGAGTGAACAGCTAAGAGCCTGCAGCCAAGCG 81384

RESULT 3
AF217974 2070 bp mRNA linear PRI 01-OCT-2000
LOCUS Homo sapiens clone PPI396 unknown mRNA.
DEFINITION AF217974
ACCESSION AF217974
VERSION AF217974.1 GI:10441878
KEYWORDS FLI CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
AUTHORS Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu.Y.,
Yu,J. and Han,L.H.
TITLE Novel Human CDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2070)
AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,
Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu.Y.,
Yu,J. and Han,L.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related
genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-Tu Road, Shanghai
200032, People's Republic of China
FEATURES
source
1..2070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PPI396"
602..1315
/codon_start=1
/product="unknown"
/protein_id="AAG17217.1"
/db_xref="GI:10441879"
/translaton="MGCTTATFHSDDSSKSCACGGGIORRSVCLGSAALPGGG
EAGAGTGSCTGSRPDRACSLGPCERTWRWTGTGPGSCSCEGSGTORDILCVS
KLSTFNVTSPNSCHLPRPALQPCQACQDRWSTPMSRCSRCSGGTQREVQC
LSTNLTSTPCPLRFSRFRKPNCSQPCSRPDDQCKDSSPHCLPLVVOARLVPIYT
ATCCRSCHVLRSPQDPS"
BASE COUNT 409 a 670 c 585 g 406 t
ORIGIN

Query Match 20.1%; Score 529; DB 9; Length 2070;
Best Local Similarity 99.8%; Pred. No. 1.7e-286;
Matches 649; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1910 ggacccagggcaccctccagcgtcaggtgcggtatccccagatgcgcgcgcgcgcgcac 1969
Db 1 GGACCCAGGCACCTCCAGCGTCAGTGCAGTCCCCAGATGCCGCCGCCGCCATC 60

QY 1970 ccagagacccctggggtctccagctgcgtactggaacagagtgaggacactctgcagtct 2029
Db 61 CCAGAGACACCCCTGGGGTCTCCAGCTGGCTACTGGAACAGAGTGGGACACTCTGCATGCT 120

QY 2030 cagcgtctcgggaaagtctctggcccccatttctctgcatctcccgtagtcgg 2089
Db 121 CAGCGTCTCTGGGGAAGAGTGTCTGGCCCCCATTTTCTCTGTGATCTCCCGTGAAGTCGG 180

QY 2090 gagaggaactggatgaacgacagctgtgcgggggtgccaggcccccacgctccctgaac 2149

```

```

Db 181 GAGAGGAACCTGGATGAACGACAGCTGTGCCCGGGTGCCAGGCCCCAGCCTCCCTGAAC 240
QY 2150 cctgcacagcaccatccccccatactggagagtgagcagtgagacatcctgcagcc 2209
Db 241 CCTG-CACGGCACCCCATGCCCCATAGTGGAGGCTGGCGAGTGGACATCCTGCAGCC 299
QY 2210 gctcctgtggccccggcaccacagcaccgacagtcagtcgagtcgaggaagtaattggggggg 2269
Db 300 GTCCTCTGTGCCCGCGCACCCAGCACCGGACAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 359
QY 2270 gtgctcctcgggtgccccggagcgtgtggacatctccccggcgccccacatcacccag 2329
Db 360 GTGCTCTCTGGTGCCTCCCGGAGCGCTGTGGACATCTCCCGCGGCCACACATCACCACT 419
QY 2330 cttgcagctgcgctctgtggccattgggaagtgtgctctccttgagcagtgctcg 2389
Db 420 CTTGCCAGCTGCCTCTGTGGCCATTGGAGTGGCTCTCCTTGGAGCAGTGTCTCG 479
QY 2390 tgcggtgcggcgccgacgaagccgaggttcgctgtgtgtgggaaacagctgatg 2449
Db 480 TGGGTGCGCGCGCGCGGCGAGAGCGCGAGGTTCGCTGTGTGGGAACACGGTGTG 539
QY 2450 aagtgcagcagcagagtgctgcagggccccccacagccccccagcagagagggctgtg 2509
Db 540 AAGTGAGCGAGCAGGAGTGTGCTCAGGCCGCCACAGCCCCCAGCAGAGAGGCGCTGTG 599
QY 2510 acatggggccttactactgctgttccacagcagctggagctccaag 2559
Db 600 ACATGGGGCCTTACTACTGCTGTGCTCAGCGGACTGGAGCTCCAAG 649

RESULT 4
AC053497
LOCUS 207815 bp DNA linear HTG 13-AUG-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
AC053497
ACCESSION AC053497
VERSION AC053497.4 GI:9797901
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Waterston,R.H.
AUTHORS The sequence of Homo sapiens clone
TITLE Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 207815)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 12, 2000 this sequence version replaced gi:8439979.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0243G22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-ff
Insert size: 204215; sum-of-contigs
Quality coverage: 6.36 in Q20 bases; agarose-ff

```



```
AC053497
AC053497.4 GI:9797901
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 207815)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 207815)
Waterston,R.H.
Direct Submission
Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2000 this sequence version replaced gi:8439979.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0243Q22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-ip
Insert size: 204215; sum-of-contigs
Quality coverage: 6.36 in Q20 bases; agarose-fp
Quality coverage: 6.31 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1381: contig of 1381 bp in length
1382 1481: gap of unknown length
1482 2602: contig of 1121 bp in length
2603 2702: gap of unknown length
2703 4013: contig of 1311 bp in length
4014 4113: gap of unknown length
4114 5331: contig of 1218 bp in length
5332 5431: gap of unknown length
5432 6622: contig of 1191 bp in length
6623 8166: contig of 1444 bp in length
8167 9674: gap of unknown length
9675 10943: contig of 1408 bp in length
10944 11044: gap of unknown length
11045 12198: contig of 1155 bp in length
12199 14193: gap of unknown length
14194 14293: contig of 1895 bp in length
14294 15717: gap of unknown length
15718 15817: contig of 1424 bp in length
15818 16935: gap of unknown length
16936 17035: contig of 1118 bp in length
17036 18375: gap of unknown length
18376 18475: contig of 1340 bp in length
18476 18475: gap of unknown length

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

vector_side:left"
misc_feature 45874..55056 /note="assembly_name:Contig70"
misc_feature 55157..67059 /note="assembly_name:Contig71"
misc_feature 67160..80260 /note="assembly_name:Contig72"
misc_feature 80361..96528 /note="assembly_name:Contig73"
misc_feature 96629..114529 /note="assembly_name:Contig74"
misc_feature 114630..133673 /note="assembly_name:Contig75"
misc_feature 133774..153791

Query Match 17.4%; Score 457; DB 2; Length 207815;
Best Local Similarity 99.7%; Pred. No. 7.9e-246;
Matches 697; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 433 aggtccggcttcgagaccatcaagcaggaagtgtcggttatgtggagagtgcccttt 492
Db 65439 AGGTCCGGCTTCGAGACCCATCAAGCCAGAGTGTTCGGTTATGGGAGAGTGCCTTT 65498
QY 493 gcattgcactgcaccggaacccagcagccctcgagagccaccagatctgagctgtcc 552
Db 65499 GCATTGCCACTGCACCGGAACCCGAGCACCCTCGGAGGCCACCACAGATCTGAGCTGTC 65558
QY 553 ctgctctcttcagagggagaggtcttcctccctactccaagagcagagccattc 612
Db 65559 CTGATCTCTTCTAGAGGGGAGAGGCTTATTCGGTCCCTACTCCAGAGAGAGCCATTC 65618
QY 613 tccgcaaacgagcagcccccacaaactgagctccctcccaagaaatgtctgtccacaccca 672
Db 65619 TCCGCAAAACGAGCAGCCCAAACTGAGCTCCCTCCCAAGAACTGTGTGTCACACCCCA 65678
QY 673 tcccccacagcaaaccttaagccctgaactgctcagacagagtgcccccagaacc 732
Db 65679 TCCCCCAAGCAGAACCTCTAAGCCCTGAACCTGCTCGACAGAGGTGGCCCCCAGAAC 65738
QY 733 aggcctgcacctacggcatcacccagagccagcagcctcttgccacagagcccccctca 792
Db 65739 AGGCCTG-CCCTCTACGGCATCACCCAGAGCCCGAGGCCCTCTGSCACAGAGCCCTCTCA 65797
QY 793 ccaagcactcttaggaagagtggttttccttccttccttccttccttccttccttccttc 852
Db 65798 CCCACGCACTCCTTAGGAGAGGTGGGTCTTTCCTGTCATCCCTCCAGCCACGAGGCCCA 65857
QY 853 agttccaggggttgccagtcctccaggtagcagggagagcgcctgatccttttccttcg 912
Db 65858 AGTTCCAGGGTTCGGCCAGTCCCTCAGTACGAGGAGAGCCCTGATCCTTTTCCTTCG 65917
QY 913 gtccctcgggcccagggccagcagggcccttcgggaaaggggggggactcctcac 972
Db 65918 GTCCCTCGGGCCCGAGCCAGAGGG-CAAGGGGCTTTGGGGAACGGGGGAGCTCCTCAC 65976
QY 973 gggcccccgcctggagcctgacccctcagcaccggcgccgtggtgcccctgctgagcaac 1032
Db 65977 GGGCCCCGCTTGGAGCTTGACCTCAGCACCCGGCGGCTGGCTGCCCTCTCTGAGCAAC 66036
QY 1033 gggcccccgcagctcctctgagcctcttttctccagtagcctattcccaagatgt 1092
Db 66037 GGCCCCCATGCGAGTCCCTCTGGAGCCTCTTGTCTCCAGTAGCCCTATTCCAGATGT 66096
QY 1093 tctggggagagtgaaacagcctaagacctgcagccaagcg 1131
Db 66097 TCTGGGAGAGTGAAACAGCTAAGAGCTTGACAGCTTGACCAAGCG 66135

RESULT 5
AC053497/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
```

```

* 18476 19962: contig of 1487 bp in length
* 19963 20062: gap of unknown length
* 20063 21169: contig of 1107 bp in length
* 21170 21269: gap of unknown length
* 21270 22452: contig of 1183 bp in length
* 22453 22552: gap of unknown length
* 22553 24582: contig of 2030 bp in length
* 24583 24683: gap of unknown length
* 24684 26483: contig of 1701 bp in length
* 26484 27110: contig of 1227 bp in length
* 27111 27810: gap of unknown length
* 27811 28195: contig of 1385 bp in length
* 28196 29295: gap of unknown length
* 29296 30800: contig of 1505 bp in length
* 30801 30900: gap of unknown length
* 30901 32555: contig of 1655 bp in length
* 32556 34653: gap of unknown length
* 34654 34753: gap of unknown length
* 34754 37122: contig of 2369 bp in length
* 37123 37222: gap of unknown length
* 37223 39506: contig of 2284 bp in length
* 39507 39607: gap of unknown length
* 39608 41220: contig of 1614 bp in length
* 41221 41320: gap of unknown length
* 41321 42378: contig of 1658 bp in length
* 42379 43078: gap of unknown length
* 43079 45773: contig of 2895 bp in length
* 45774 50556: contig of 9183 bp in length
* 50557 55156: gap of unknown length
* 55157 67059: contig of 11903 bp in length
* 67060 67159: gap of unknown length
* 67160 80260: contig of 13101 bp in length
* 80261 80360: gap of unknown length
* 80361 96529: contig of 16168 bp in length
* 96530 96629: gap of unknown length
* 96630 114530: contig of 17901 bp in length
* 114531 133673: gap of unknown length
* 133674 133773: gap of unknown length
* 133774 153791: contig of 20018 bp in length
* 153792 153891: gap of unknown length
* 153892 176079: contig of 22188 bp in length
* 176080 207815: gap of unknown length
* 176180 207815: contig of 31636 bp in length.

```

## FEATURES

source

```

1..207815 Homo sapiens
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-243G22"
1..1381
/note="assembly_name:Contig13"
1482..2602
/note="assembly_name:Contig15"
2703..4013
/note="assembly_name:Contig21"
4114..5331
/note="assembly_name:Contig25"
5432..6622
/note="assembly_name:Contig33"
6723..8166
/note="assembly_name:Contig37"
8267..9674
/note="assembly_name:Contig38"
9775..10943
/note="assembly_name:Contig40"
11044..12198
/note="assembly_name:Contig42"
12299..14193
/note="assembly_name:Contig45"

```

```

misc_feature 14294..15717
/note="assembly_name:Contig47"
misc_feature 15818..16935
/note="assembly_name:Contig50"
misc_feature 17036..18375
/note="assembly_name:Contig51"
misc_feature 18476..19962
/note="assembly_name:Contig52"
misc_feature 20063..21169
/note="assembly_name:Contig53"
misc_feature 21270..22452
/note="assembly_name:Contig54"
misc_feature 22553..24582
/note="assembly_name:Contig56"
misc_feature 24683..26383
/note="assembly_name:Contig57"
misc_feature 26484..27710
/note="assembly_name:Contig58"
misc_feature 27811..29195
/note="assembly_name:Contig59"
misc_feature 29296..30800
/note="assembly_name:Contig60"
misc_feature 30901..32555
/note="assembly_name:Contig63"
misc_feature 32656..34653
/note="assembly_name:Contig64"
misc_feature 34754..37122
/note="assembly_name:Contig65"
misc_feature 37223..39506
/note="assembly_name:Contig66"
misc_feature 39607..41220
/note="assembly_name:Contig67"
misc_feature 41321..42978
/note="assembly_name:Contig68"
misc_feature 43079..45773
/note="assembly_name:Contig69"
clone_end:T7
vector_side:left
45874..55056
/note="assembly_name:Contig70"
55157..67059
/note="assembly_name:Contig71"
67160..80260
/note="assembly_name:Contig72"
80361..96528
/note="assembly_name:Contig73"
96629..114529
/note="assembly_name:Contig74"
114630..133673
/note="assembly_name:Contig75"
133774..153791

```

Query Match 7.8%; Score 205; DB 2; Length 207815;

Best Local Similarity 100.0%; Pred. No. 1.5e-103;

Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2178 ctggagactgagtgagatcctgcagcgcctctgtggcccgccagcaccg 2237

DB 131101 CTGGAGGCTGCGAGTGGACATCTCTGAGCGGCTCTGTGGCCCGGACCCAGCAG 131042

QY 2238 ccagctgcagtcgagcaggaatttgggggggtggtcctcctggtgcctccgagcgtg 2297

DB 131041 CCAGCTGCAGTCCCGCAGGAATTTGGGGGGGTGGCTCCTCGTGCCTCCGAGCGCTG 130982

QY 2298 tggacatctccccggcccaacatcaccagctcttgccagcttgccctctgtgccattg 2357

DB 130981 TGGACATCTCCCCGGCCCAACATCACCAGCTTGGCCAGCTGCGCTCTGTGGCCATTG 130922

QY 2358 ggaagtgtgctctccttgagccag 2382

DB 130921 GGAAGTTGGCTCTCCTTGGAGCCAG 130897

```
RESULT 6
AC092479 142023 bp DNA linear HTG 05-FEB-2002
LOCUS Mus musculus clone rp23-231115 strain C57BL/6J, WORKING DRAFT
DEFINITION SEQUENCE, 3 unordered pieces.
ACCESSION AC092479.12 GI:18497125
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 142023)
AUTHORS Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-231115
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142023)
AUTHORS Jia,H., Zhang,P., Lin,S., Wu,H. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Feb 5, 2002 this sequence version replaced gi:18390246.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 32949: contig of 32949 bp in length
* 32950 33049: gap of unknown length
* 78918 79018: contig of 45869 bp in length
* 78919 79018: gap of unknown length
* 79019 142023: contig of 63005 bp in length.
FEATURES
source
1..142023
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="rp23-231115"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 37523 a 34071 c 31510 g 38713 t 206 others
ORIGIN
Query Match 1.5%; Score 39; DB 2; Length 142023;
Best Local Similarity 100.0%; Pred. No. 7.6e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2290 gagcgctgtggacatctccccggcccaacatcaccag 2328
|||||
Db 86350 GAGCGTGTGGACATCTCCCCGGCCCAACATCACCAG 86388
|||||

RESULT 7
AC093317 215810 bp DNA linear HTG 29-JAN-2002
LOCUS Mus musculus clone rp23-218k6 strain C57BL/6J, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.
ACCESSION AC093317
VERSION AC093317.16 GI:18390243
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 215810)
AUTHORS Jia,H., Zhang,P., Lin,S. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-218k6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215810)
AUTHORS Jia,H., Zhang,P., Lin,S. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jan 29, 2002 this sequence version replaced gi:18139414.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2528: contig of 2528 bp in length
* 2529 2629: gap of unknown length
* 8258 8358: gap of unknown length
* 8358 14071: contig of 5714 bp in length
* 14072 14171: gap of unknown length
* 14172 28167: contig of 13996 bp in length
* 28168 28267: gap of unknown length
* 28268 41090: contig of 12823 bp in length
* 41091 41190: gap of unknown length
* 41191 58658: contig of 17468 bp in length
* 58659 58759: gap of unknown length
* 58759 83163: contig of 24405 bp in length
* 83164 83264: gap of unknown length
* 83264 107370: contig of 24107 bp in length
* 107371 137048: contig of 29578 bp in length
* 137049 137149: gap of unknown length
* 137149 215810: contig of 78662 bp in length.
FEATURES
source
1..215810
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="rp23-218k6"
/clone_lib="RPCI mouse BAC library 23"
BASE COUNT 59907 a 50405 c 47367 g 57018 t 1113 others
ORIGIN
Query Match 1.5%; Score 39; DB 2; Length 215810;
Best Local Similarity 100.0%; Pred. No. 7.6e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2290 gagcgctgtggacatctccccggcccaacatcaccag 2328
|||||
Db 114431 GAGCGTGTGGACATCTCCCCGGCCCAACATCACCAG 114469
|||||

RESULT 8
BC016215 1409 bp mRNA linear ROD 05-NOV-2001
LOCUS Mus musculus, Similar to hypothetical protein FLJ13710, clone
DEFINITION MGC:28749 IMAGE:4482484, mRNA, complete cds.
ACCESSION BC016215
VERSION BC016215.1 GI:16740678
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
```

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE	AUTHORS	TITLE
1	...	...
2	...	...
3	...	...
4	...	...
5	...	...
6	...	...
7	...	...
8	...	...
9	...	...
10	...	...
11	...	...
12	...	...
13	...	...
14	...	...
15	...	...
16	...	...
17	...	...
18	...	...
19	...	...
20	...	...
21	...	...
22	...	...
23	...	...
24	...	...
25	...	...
26	...	...
27	...	...
28	...	...
29	...	...
30	...	...
31	...	...
32	...	...
33	...	...
34	...	...
35	...	...
36	...	...
37	...	...
38	...	...
39	...	...
40	...	...
41	...	...
42	...	...
43	...	...
44	...	...
45	...	...
46	...	...
47	...	...
48	...	...
49	...	...
50	...	...
51	...	...
52	...	...
53	...	...
54	...	...
55	...	...
56	...	...
57	...	...
58	...	...
59	...	...
60	...	...
61	...	...
62	...	...
63	...	...
64	...	...
65	...	...
66	...	...
67	...	...
68	...	...
69	...	...
70	...	...
71	...	...
72	...	...
73	...	...
74	...	...
75	...	...
76	...	...
77	...	...
78	...	...
79	...	...
80	...	...
81	...	...
82	...	...
83	...	...
84	...	...
85	...	...
86	...	...
87	...	...
88	...	...
89	...	...
90	...	...
91	...	...
92	...	...
93	...	...
94	...	...
95	...	...
96	...	...
97	...	...
98	...	...
99	...	...
100	...	...

**10**

REMARK

COMMENT

## FEATURES

source

CDS

BASE COUNT	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

ORIGIN

Query M

### Best Lo Matches

Qy 1600

Db 313



mus musculus clone RP23-176010, LOW-PASS SEQUENCE SAMPLING.

AC101188

AC101188.1 GI:17059962

HTG; HTGS-PHASE0.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 62385)

Barren,B., Linton,L., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-176010

Unpublished

2 (bases 1 to 62385)

Barren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campatiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeLallano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meidrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Rettar,R., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zemбек,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information -----

Center project name: L16125

Center clone name: 176\_Q10

-----

\* NOTE: This record contains 80 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1 692: contig of 692 bp in length

\* 693 792: gap of 100 bp

\* 793 1503: contig of 711 bp in length

\* 1504 1603: gap of 100 bp

\* 1604 2312: contig of 709 bp in length

\* 2313 2412: gap of 100 bp

\* 2413 3066: contig of 656 bp in length

\* 3069 3168: gap of 100 bp

\* 3169 3846: contig of 678 bp in length

\* 3847 3946: gap of 100 bp

\* 3947 4630: contig of 684 bp in length

\* 4631 4730: gap of 100 bp

\* \*

\* 4731 5409: contig of 679 bp in length  
\* 5410 5509: gap of 100 bp  
\* 5510 6217: contig of 708 bp in length  
\* 6218 6317: gap of 100 bp  
\* 6318 6972: contig of 655 bp in length  
\* 6973 7072: gap of 100 bp  
\* 7073 7767: contig of 695 bp in length  
\* 7768 7867: gap of 100 bp  
\* 7868 8572: contig of 705 bp in length  
\* 8573 8672: gap of 100 bp  
\* 8673 9377: contig of 705 bp in length  
\* 9378 9477: gap of 100 bp  
\* 9478 10144: contig of 667 bp in length  
\* 10145 10244: gap of 100 bp  
\* 10245 10883: contig of 639 bp in length  
\* 10884 10983: gap of 100 bp  
\* 10984 11665: contig of 682 bp in length  
\* 11666 11765: gap of 100 bp  
\* 11766 12433: contig of 668 bp in length  
\* 12434 12533: gap of 100 bp  
\* 12534 13214: contig of 681 bp in length  
\* 13215 13314: gap of 100 bp  
\* 13315 14007: contig of 693 bp in length  
\* 14008 14107: gap of 100 bp  
\* 14108 14811: contig of 704 bp in length  
\* 14812 14911: gap of 100 bp  
\* 14912 15614: contig of 703 bp in length  
\* 15615 15714: gap of 100 bp  
\* 15715 16428: contig of 714 bp in length  
\* 16429 16528: gap of 100 bp  
\* 16529 17220: contig of 692 bp in length  
\* 17221 17320: gap of 100 bp  
\* 17321 17979: contig of 659 bp in length  
\* 17980 18079: gap of 100 bp  
\* 18080 18759: contig of 680 bp in length  
\* 18760 18859: gap of 100 bp  
\* 18860 19526: contig of 667 bp in length  
\* 19527 19626: gap of 100 bp  
\* 19627 20303: contig of 677 bp in length  
\* 20304 20403: gap of 100 bp  
\* 20404 21090: contig of 687 bp in length  
\* 21091 21190: gap of 100 bp  
\* 21191 21897: contig of 707 bp in length  
\* 21898 21997: gap of 100 bp  
\* 21998 22668: contig of 671 bp in length  
\* 22669 22768: gap of 100 bp  
\* 22769 23467: contig of 699 bp in length  
\* 23468 23567: gap of 100 bp  
\* 23568 24257: contig of 690 bp in length  
\* 24258 24357: gap of 100 bp  
\* 24358 25045: contig of 688 bp in length  
\* 25046 25145: gap of 100 bp  
\* 25146 25810: contig of 665 bp in length  
\* 25811 25910: gap of 100 bp  
\* 25911 26573: contig of 663 bp in length  
\* 26574 26673: gap of 100 bp  
\* 26674 27341: contig of 668 bp in length  
\* 27342 27441: gap of 100 bp  
\* 27442 28139: contig of 688 bp in length  
\* 28130 28229: gap of 100 bp  
\* 28230 28907: contig of 678 bp in length  
\* 28908 29007: gap of 100 bp  
\* 29008 29697: contig of 690 bp in length  
\* 29698 29797: gap of 100 bp  
\* 29798 30503: contig of 706 bp in length  
\* 30504 30603: gap of 100 bp  
\* 30604 31303: contig of 700 bp in length  
\* 31304 31403: gap of 100 bp  
\* 31404 32047: contig of 644 bp in length  
\* 32048 32147: gap of 100 bp  
\* 32148 32833: contig of 686 bp in length  
\* 32834 32933: gap of 100 bp  
\* 32934 33615: contig of 682 bp in length

\* 33616 33715: gap of 100 bp  
\* 33716 34416: contig of 701 bp in length  
\* 34417 34516: gap of 100 bp  
\* 34517 35189: contig of 673 bp in length  
\* 35190 35289: gap of 100 bp  
\* 35290 35952: contig of 663 bp in length  
\* 35953 36052: gap of 100 bp  
\* 36053 36710: contig of 658 bp in length  
\* 36711 36810: gap of 100 bp  
\* 36811 37486: contig of 676 bp in length  
\* 37487 37586: gap of 100 bp  
\* 37587 38276: contig of 690 bp in length  
\* 38277 38376: gap of 100 bp  
\* 38377 39089: contig of 713 bp in length  
\* 39090 39189: gap of 100 bp  
\* 39190 39825: contig of 636 bp in length  
\* 39826 39925: gap of 100 bp  
\* 39926 40531: contig of 606 bp in length  
\* 40532 40631: gap of 100 bp  
\* 40632 41161: contig of 530 bp in length  
\* 41162 41261: gap of 100 bp  
\* 41262 41931: contig of 670 bp in length  
\* 41932 42031: gap of 100 bp  
\* 42032 42699: contig of 668 bp in length  
\* 42700 42799: gap of 100 bp  
\* 42800 43474: contig of 675 bp in length  
\* 43475 43574: gap of 100 bp  
\* 43575 44272: contig of 698 bp in length  
\* 44273 44372: gap of 100 bp  
\* 44373 45069: contig of 697 bp in length  
\* 45070 45169: gap of 100 bp  
\* 45170 45836: contig of 667 bp in length  
\* 45837 45936: gap of 100 bp  
\* 45937 46642: contig of 705 bp in length  
\* 46642 46741: gap of 100 bp  
\* 46742 47447: contig of 706 bp in length  
\* 47448 47547: gap of 100 bp  
\* 47548 48249: contig of 702 bp in length  
\* 48250 48349: gap of 100 bp  
\* 48350 49000: contig of 651 bp in length  
\* 49001 49100: gap of 100 bp  
\* 49101 49777: contig of 677 bp in length  
\* 49778 49877: gap of 100 bp  
\* 49878 50555: contig of 678 bp in length  
\* 50556 50655: gap of 100 bp  
\* 50656 51344: contig of 689 bp in length  
\* 51345 51444: gap of 100 bp  
\* 51445 52140: contig of 696 bp in length  
\* 52141 52240: gap of 100 bp  
\* 52241 52926: contig of 686 bp in length  
\* 52927 53026: gap of 100 bp  
\* 53027 53732: contig of 706 bp in length  
\* 53733 53832: gap of 100 bp

Query Match 0.9%; Score 23; DB 2; Length 62385;  
Best Local Similarity 100.0%; Pred.No. 0.81;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 ccagccctgcgggtgggggtg 192  
|||||  
Db 51295 CCACGCCCTGCGGGTGGGGGTG 51273

RESULT 12  
AC026636/c  
LOCUS AC026636 151333 bp DNA linear HTG 27-AUG-2001  
DEFINITION Homo sapiens chromosome 15 clone RP11-170E16 map 15, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 4 unordered pieces.  
ACCESSION AC026636  
VERSION AC026636.5 GI:5291088  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 151333)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-170E16
Unpublished
2 (bases 1 to 151333)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukagaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Horton,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Jones,C., Kann,L., Karatas,A.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Lander,E., Lander,L.,
Klein,J., LaRoque,K., Lamazares,K., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N., Rothman,D.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Spencer,B.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spence,R., Talamas,J.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 26, 2001 this sequence version replaced gi:15148317.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8704
Center clone name: 170_E_16
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 25339: contig of 25339 bp in length
* 25340 25439: gap of 100 bp
* 25440 36562: contig of 11123 bp in length
* 36563 36662: gap of 100 bp
* 36663 61806: contig of 25144 bp in length
* 61807 61906: gap of 100 bp
* 61907 151333: contig of 89427 bp in length.
FEATURES
source
Location/Qualifiers
1. .151333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-170E16"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 44563 a 32212 c 31615 g 42618 t 325 others
ORIGIN
Query Match 0.98; Score 23; DB 2; Length 151333;
Best Local Similarity 100.0%; Pred. No. 0.82;

```

```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1600 tccatcatcaatgggaactgggc 1622
|||||
Db 32392 tccatcatcaatgggaactgggc 32370
|||||
RESULT 13
AC104938/c
LOCUS
DEFINITION
Homo sapiens chromosome 15 clone CTB-2524L6 map 15, WORKING DRAFT
SEQUENCE, 7 ordered pieces.
AC104938
AC104938.1 GI:17977316
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184483)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone CTB-2524L6
Unpublished
2 (bases 1 to 184483)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukagaiter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., Levine,R., Liu,G.,
Lamazan,C., Landers,T., Lehoczy,J., Levine,R., Matthews,C.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21457
Center clone name: 2524_L_6
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 182343 bases at least Q40
Consensus quality: 182981 bases at least Q30
Consensus quality: 183229 bases at least Q20
Insert size: 183883; sum-of-contigs
Quality coverage: 10.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces

```

```

TITLE
JOURNAL
COMMENT

```

\* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* the accession number will be preserved.  
 \* 1 549: contig of 549 bp in length  
 \* 550 649: gap of 100 bp  
 \* 650 14177: contig of 13528 bp in length  
 \* 14178 14277: gap of 100 bp  
 \* 14278 22037: contig of 7760 bp in length  
 \* 22038 22137: gap of 100 bp  
 \* 22138 40150: contig of 18013 bp in length  
 \* 40151 40250: gap of 100 bp  
 \* 40251 66123: contig of 25873 bp in length  
 \* 66124 66223: gap of 100 bp  
 \* 66224 114326: contig of 48103 bp in length  
 \* 114327 114426: gap of 100 bp  
 \* 114427 184483: contig of 70057 bp in length.

## FEATURES

```

source
1..184483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTB-2524L6"
/clone_lib="CIRB Human BAC"
1..549
/note="assembly_fragment"
550..14177
/note="assembly_fragment"
14278..22037
/note="assembly_fragment"
22138..40150
/note="assembly_fragment"
40251..66123
/note="assembly_fragment"
66224..114326
/note="assembly_fragment"
114327..184483
/note="assembly_fragment"
55127 a 39050 c 39152 g 50554 t 600 others
ORIGIN
```

Query Match 0.9%; Score 23; DB 2; Length 184483;  
 Best Local Similarity 100.0%; Pred. NO. 0.82;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1500 tccatcatcaatgggaactgggc 1622

Db 154963 TCCATCATCAATGGGAACGGC 154941

## RESULT 14

AC019036/c 168872 bp DNA linear HTG 08-APR-2000  
 LOCUS  
 DEFINITION Homo sapiens chromosome 4 clone RP11-12318 map 4, WORKING DRAFT  
 SEQUENCE, 20 unordered pieces.

AC019036  
 AC019036 GI:7528071  
 VERSION  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 168872)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## TITLE

Homo sapiens chromosome 4, clone RP11-12318

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 168872)

## AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
 Derrellano,K., Dewar,K., Domino,M., Doyle,M., Fencstor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Rothman,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu.X., Wyman,D., Ye.W.J.,  
 Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (30-DEC-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 8, 2000 this sequence version replaced g1:6730845.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4941

Center clone name: 123.I.8

----- Summary Statistics

Sequencing Vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156691 bases at least Q40

Consensus quality: 162047 bases at least Q30

Consensus quality: 164542 bases at least Q20

Insert size: 166000; agarose-gp

Insert size: 166972; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; agarose-gp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 20 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1092: contig of 1092 bp in length  
 1093 1192: gap of 100 bp  
 1193 4028: contig of 2836 bp in length  
 4029 4128: gap of 100 bp  
 4129 8354: contig of 4226 bp in length  
 8355 8454: gap of 100 bp  
 8455 11747: contig of 3293 bp in length  
 11748 11847: gap of 100 bp  
 11848 16091: contig of 4244 bp in length  
 16092 16191: gap of 100 bp  
 16192 19148: contig of 2957 bp in length  
 19149 19248: gap of 100 bp  
 19249 25597: contig of 6349 bp in length  
 25598 25697: gap of 100 bp  
 25698 31571: contig of 5874 bp in length  
 31572 31671: gap of 100 bp  
 31672 38194: contig of 6523 bp in length  
 38195 38294: gap of 100 bp  
 38295 44033: contig of 5739 bp in length  
 44034 44133: gap of 100 bp  
 44134 51560: contig of 7427 bp in length  
 51561 51660: gap of 100 bp  
 51661 60691: contig of 9031 bp in length

```

* 60592 60791: gap of 100 bp
* 60792 69942: contig of 9151 bp in length
* 69943 70042: gap of 100 bp
* 70043 77752: contig of 7710 bp in length
* 77753 77852: gap of 100 bp
* 77853 89028: contig of 11176 bp in length
* 89029 89128: gap of 100 bp
* 89129 99478: contig of 10350 bp in length
* 99479 99578: gap of 100 bp
* 99579 115864: contig of 16286 bp in length
* 115865 115964: gap of 100 bp
* 115965 134124: contig of 18160 bp in length
* 134125 134224: gap of 100 bp
* 134225 152621: contig of 18397 bp in length
* 152622 152721: gap of 100 bp
* 152722 168872: contig of 16151 bp in length.

```

## FEATURES

source

```

1. .168872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="4"
/map="4"
/clone="RP11-12318"
/clone_lib="RP11-11 Human Male BAC"
1. .1092
/note="assembly_fragment"
1193. .4028
/note="assembly_fragment"
4129. .8354
/note="assembly_fragment"
8455. .11747
/note="assembly_fragment"
11848. .16091
/note="assembly_fragment"
16192. .19148
/note="assembly_fragment"
19249. .25597
/note="assembly_fragment"
25698. .31571
/note="assembly_fragment"
31672. .38194
/note="assembly_fragment"
38295. .44033
/note="assembly_fragment"
44134. .51560
/note="assembly_fragment"
51661. .60691
/note="assembly_fragment"
60792. .69942
/note="assembly_fragment"
70043. .77752
/note="assembly_fragment"
77853. .89028
/note="assembly_fragment"
89129. .99478
/note="assembly_fragment"
99579. .115864
/note="assembly_fragment"
clone_end:SP6
vector_side:right
115965. .134124
/note="assembly_fragment"
134225. .152621
/note="assembly_fragment"
152722. .168872
/note="assembly_fragment"
BASE COUNT 52704 a 30458 c 31011 g 52799 t 1900 others
ORIGIN

```

Query Match 0.8%; Score 22; DB 2; Length 168872;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 485 tgcctttgcattgccactgca 506
|||||
DB 115680 tgcctttgcattgccactgca 115659

```

## RESULT 15

```

LOCUS AC097488 175970 bp DNA linear PRI 09-FEB-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-218C23, complete sequence.
ACCESSION AC097488 AC067828
VERSION AC097488.2 GI:18642911
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 175970)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175970)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

REFERENCE 3 (bases 1 to 175970)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

```

```

COMMENT
On Feb 9, 2002 this sequence version replaced gi:16259095.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H_NH0218C23
Drafting center: WIBR
-----

```

## FEATURES

source

```

1. .175970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-218C23"
BASE COUNT 50188 a 32373 c 33188 g 60221 t
ORIGIN

```

Query Match 0.8%; Score 22; DB 9; Length 175970;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 485 tgcctttgcattgccactgca 506
|||||
DB 133569 TGCCCTTGCATTGCCACTGCA 133590

```

Search completed: July 23, 2002, 20:20:02  
 Job time: 9842 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 13:05:10 ; Search time 3244.83 Seconds  
(without alignments)  
16987.176 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atggagaactgactgcag.....cctcagcctttccagcatag 2634

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hcg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_un.\*

28: em\_vl.\*

29: em\_hcg\_hum.\*

30: em\_hcg\_inv.\*

31: em\_hcg\_inv.\*

32: em\_hcg\_inv.\*

33: em\_hcg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	2557.4	97.1	4230	6	AX342636	AX342636 Sequence
2	878.2	33.3	176550	9	AL356356	Human DNA
3	854.2	32.4	207815	2	AC053497	Homo sapi
4	638	24.2	2070	9	AF217974	Homo sapi
5	251.4	9.5	142023	2	AC092479	Mus muscu
6	250.2	9.5	207815	2	AC053497	Homo sapi
7	239.4	9.1	215810	2	AC093317	Homo sapi
8	211.6	8.0	2703	9	BC008840	Homo sapi
9	165.2	6.3	2964	9	AK023772	Homo sapi
10	163.4	6.2	4271	10	AF314171	Mus muscu
11	113.2	4.3	2805	6	AX319854	Sequence
12	106.8	4.1	3675	6	AX319852	Sequence
13	99	3.8	2410	6	AX342638	Sequence
14	97	3.7	86446	2	AC014412	Drosophil
15	97	3.7	182525	3	AC009213	Drosophil
16	97	3.7	239171	3	AE003765	Drosophil
17	96.6	3.7	5061	6	E55265	Novel metal
18	96.2	3.7	5353	6	AX128433	Sequence
19	94	3.6	2848	9	AF140674	Homo sapi
20	93	3.5	11008	3	AF078161	Manduca s
21	92.2	3.5	7100	3	AF205357	Drosophil
22	91.6	3.5	3312	6	AX319851	Sequence
23	91.6	3.5	3312	6	E55273	Novel metal
24	91.4	3.5	3132	6	AX327754	Sequence
25	91.4	3.5	3139	6	AX111571	Sequence
26	91.4	3.5	3329	6	AX327752	Sequence
27	88.4	3.4	3400	9	AF163762	Homo sapi
28	84.8	3.4	125020	9	AF429315	Homo sapi
29	80.6	3.1	4287	6	AX319855	Sequence
30	78	3.0	125020	9	AF429315	Homo sapi
31	77.4	2.9	1119	6	AX224813	Sequence
32	77.4	2.9	1953	6	AX224817	Sequence
33	77.4	2.9	2316	6	AX224821	Sequence
34	77.4	2.9	4854	6	AX224823	Sequence
35	77.2	2.9	3179	9	AF140675	Homo sapi
36	74.2	2.8	4067	9	AB011177	Homo sapi
37	74.2	2.8	5199	3	AF043121	Haemochu
38	73.4	2.8	2445	6	AX327757	Sequence
39	73.4	2.8	3207	6	AX327747	Sequence
40	73.4	2.8	3402	6	AX338539	Sequence
41	73.4	2.8	3403	6	AX327745	Sequence
42	73.4	2.8	3403	6	AX338537	Sequence
43	73.4	2.8	3471	6	AX327755	Sequence
44	73	2.8	1341	6	AX224811	Sequence
45	73	2.8	2175	6	AX224815	Sequence

ALIGNMENTS

RESULT	1	AX342636	Sequence 33 from Patent WO0198468.	4230 bp	DNA	linear	PAT 12-JAN-2002
LOCUS	AX342636	Sequence 33 from Patent WO0198468.					
DEFINITION	AX342636	Sequence 33 from Patent WO0198468.					
ACCESSION	AX342636	Sequence 33 from Patent WO0198468.					
VERSION	AX342636.1	GI:18152033					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
AUTHORS	Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J., Tribouley, C.M., Deleane, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A., Fabalia, A., Khan, F.A., Wala, N.K., Yao, M.G., Lu, D.A., Patterson, C., Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R. Patent: WO 0198468-A 33 27-DEC-2001; Incyte Genomics, Inc. (US)						
JOURNAL	Incyte Genomics, Inc. (US)						
FEATURES	Location/Qualifiers						
source	1. 4230						
	/organism="Homo sapiens"						
	/db xref="taxon.9606"						
	/note="Incyte ID No: 7604035CB1"						



Db 2268 GGAAGAGTGTCGGCGCCCATTTTCCTCTGTCATCTCCCGTAGTGGGAGAGAACTG 2327  
Qy 2101 gatgaacacactatgcgcgggtgacagcccccagcctccctgaacctgcacgc 2160  
Db 2328 GATGAACACAGTGTGCGCGGGTGCCAGGCCCCAGCCTCCCTGACCCCTGCACGGC 2387  
Qy 2161 acccatgccccccatactggaggtggcggagtggaacatcctgcagccgctcctgtggc 2220  
Db 2388 ACCCATGCCCCCATACTGGAGGCTGGCGAGTGGACATCTCAGCGCGCTCCTGTGGC 2447  
Qy 2221 cccggcaccagcaccgacagctgcagtcgcggcaggaattggggggtgctctcg 2280  
Db 2448 CCGGACCCACAGCACCGCAGTGCAGTGCAGGCGCGGAGGAATTTGGGGGGGTGCTCTCG 2507  
Qy 2281 gtgccccggagcgtgtggacatctcccccggcccaacatcacccagcttggcagctg 2340  
Db 2508 GTGCCCCGGAGCGCTGTGGACATCTCCCGCGCCCAACATCACCGAGCTTGCACAGTG 2567  
Qy 2341 cgcctcttgccattgggaagtgtgctctccttggagccagtgctcogtgcggtggc 2400  
Db 2568 CGCCTCTGTGCCATTGGGAAGTTGGCTCTCCTTTGGAGCCAGTGTCTCGTGGCGGC 2627  
Qy 2401 cggggccagagaagccgcaggttcgtgtgtgggaacacggtgatgaagtgcagcag 2460  
Db 2628 CCGGCCACAGAAAGCCGACAGTTCGCTGTGTGGGAACAACGGTGTGAGTGACGCGAG 2687  
Qy 2461 caggagtgctgcagggccccccagccccccagcagagagcgtgtgacatggggccc 2520  
Db 2688 CAGGAGTGTGGCTACAGGCCCCCCACAGCCCCCAGAGAGAGGCTGTGACATGGGGCCC 2747  
Qy 2521 ttactactgctgttcacagcagactggagctccag 2559  
Db 2748 TGTACTACTGCTGTTCACAGCGACTGGAGCTCCAAG 2786

RESULT 2  
LOCUS AL356356 176550 bp DNA linear PRI 30-JAN-2002  
DEFINITION Human DNA sequence from clone Rp11-54A4 on chromosome 1, complete sequence.  
ACCESSION AL356356  
VERSION AL356356.17 GI:18476570  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 176550)  
Direct Submission  
Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Feb 1, 2002 this sequence version replaced gi:18250748.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone constructs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>  
Rp11-54A4 is from the library RpCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACE3.6  
IMPORTANT: This sequence is not the entire insert of clone Rp11-54A4. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true left end of clone Rp11-54A4 is at 1 in this sequence. The true left end of clone Rp11-363122 is at 174551 in this sequence.  
FEATURES  
Location/Qualifiers  
1..176550  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_lib="RpCI-11.1"  
misc\_feature 50307..50323  
/note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."  
misc\_feature 123022..123038  
/note="Sequence from uni-directional primer reads and dGTP big dye terminator reads only."  
BASE COUNT 45999 a 42218 c 42262 g 46071 t  
ORIGIN  
Query Match 33.3%; Score 878.2; DB 9; Length 176550;  
Best Local Similarity 84.3%; Pred. No. 2.6e-173;  
Matches 1077; Conservative 0; Mismatches 28; Indels 172; Gaps 1;  
Qy 53 tccctcagctctgcttgatcagagaggtgtgttcggagacactctcttcagacacctacag 112  
Db 80134 TCTCTCTGACCTTTTCTCTATAGTGTGTTCGGACACTCTCTTCAGACACTACAG 80193  
Qy 113 agagagccagggcccgaaaggtgtctgggaacctgtggtccagtggtgctcttgcctcc 172  
Db 80194 AGGAGGCCAGGGCCCGAAGGTGTCTGGGACCTTGGGTCCAGTGGCCCTCTGTCTCC 80253  
Qy 173 agccctgcggggtgggggtgcagcgcagggccggagacatgtcagctccctacagtgacg 232  
Db 80254 AGCCCTGCGGGTGGGGTGCAGCGCAGAGACCGGACATGTCAGCTCCCTACAGTGCGAG 80313  
Qy 233 tccaccagctctccctccctccctcccgcccccaagacatccagagccctccctccccc 292  
Db 80314 TCCACCGAGTCTGCCCTCCCTCCCGCCCCCAAGACATCCAGAGCCCTCTCTCCCCC 80373  
Qy 293 ggggccaggggtccagaccccccagacttctccagaaacctccctctgtacagagacacagt 352  
Db 80374 GGGCCAGGGTCCACAGACCCAGACTTCTCCAGAAACCTCCCTTGTACAGGACACAGT 80433  
Qy 353 ctggggaaggggtggccaccttcagagtcgcgcttccacctagggagagagagaccc 412  
Db 80434 CTCGGGAAGGGGTGGCCACTTCGAGTCCCGCTCCCACTTAGGAGAGAGAGAGACC 80493  
Qy 413 agagattcagcggccagg ----- 432  
Db 80494 AGGAGATTCGAGCGGCCAGGAGGTGAGAGGCTGGGTGGAAGAGTGGCTTGGCAAG 80553  
Qy 433 ----- 432  
Db 80554 GTGGGGCTTAGCATGAAGGAAGGGGAGCAGGGTGGCTTCAGAGGGCTTGTCTTTGGT 80613  
Qy 433 ----- 432  
Db 80614 GCCAGAGTAACACTGGGAACAGGGTGCAGAGTGTGTCAGAGCTGTGCCCCACATCTG 80673  
Qy 433 ----- aggtcccggtcttcgagaccccatcaagccaggaatgttcggttatggg 480  
Db 80674 AAACACCTTCTCAGGTCCCGCTTCGAGACCCCATCAAGCCAGCAATGTTTCGGTTATGGG 80733





**Location/Qualifiers**

```
feature
1482: 2602
/note="assembly_name:Contig13"
```

100

Country	1950	1960	1970	1980	1990	2000	2010	2020	2030	2040	2050
Japan	7	10	13	16	19	22	25	28	30	30	30
Germany	10	12	14	16	18	20	22	24	26	28	30
France	12	14	16	18	20	22	24	26	28	30	30
Italy	14	16	18	20	22	24	26	28	30	30	30
Spain	16	18	20	22	24	26	28	30	30	30	30
United Kingdom	18	20	22	24	26	28	30	30	30	30	30
Sweden	20	22	24	26	28	30	30	30	30	30	30
United States	22	24	26	28	30	30	30	30	30	30	30

[illegible]

04087 1CTGTC1GACC11111CTCTCTATATATAGG1G11G1CCGGGCGC

II3 aggaggccaggcccccgaaggctgtctgggacccctgggc

64947 AGGAGGGCCAGGGCCCCCGAAGGTGTCTGGGGACCTTGGGT

173 agccctgcggggtgggggtgcagcgaggagccggacatg

65007 AGCCCTGCGGGTGGGGTGCAGCGCAGGAGCCGGACATGT

233 tccacccgagtctgcccctccctcccccccccaagacat

65067 TCCACCCGAGTCCTCCCTCCGCGCCCAAGACAT

[illegible]

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

333 cccggggaaggggccggcccccggggcccccga

65187 CTCGGGAAGGGGTGGCCCACTTCGAGGTCCCGCTTCCCA

413 aggagattcgagcgccagg-----

65247 AGGAGATTGAGCGGCCAGGAGGTGAGAGGCCTGGGTGGA

433 -----

65307 GTGGGGGCTTAGCATGAAGGAAGGGGAGCACGGGTGGCT

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z





```

KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 207815)
              Waterston,R.H.
              The sequence of Homo sapiens clone
              2 (bases 1 to 207815)
              Waterston,R.H.
              Direct Submission
              Submitted (16-APR-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
              On Aug 12, 2000 this sequence version replaced gi:8439979.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0243G22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192323 bases at least Q40
Consensus quality: 196320 bases at least Q30
Consensus quality: 198661 bases at least Q20
Insert size: 204000; agarose-fp
Insert coverage: 6.36 in Q20 bases; agarose-fp
Quality coverage: 6.31 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1381: contig of 1381 bp in length
* 1382 1481: gap of unknown length
* 1482 2602: contig of 1121 bp in length
* 2603 2702: gap of unknown length
* 2703 4013: contig of 1311 bp in length
* 4014 4113: gap of unknown length
* 4114 5331: contig of 1218 bp in length
* 5332 5431: gap of unknown length
* 5432 6622: contig of 1191 bp in length
* 6623 6722: gap of unknown length
* 6723 8166: contig of 1444 bp in length
* 8167 8266: gap of unknown length
* 8267 9674: contig of 1408 bp in length
* 9675 9775: gap of unknown length
* 9776 10943: contig of 1169 bp in length
* 10944 11043: gap of unknown length
* 11044 12198: contig of 1155 bp in length
* 12199 12298: gap of unknown length
* 12299 14193: contig of 1895 bp in length
* 14194 14293: gap of unknown length
* 14294 15717: contig of 1424 bp in length
* 15718 15818: gap of unknown length
* 15819 16935: contig of 1118 bp in length
* 16936 17035: gap of unknown length
* 17036 18375: contig of 1340 bp in length
* 18376 18475: gap of unknown length
* 18476 19962: contig of 1487 bp in length
* 19963 20062: gap of unknown length
*
* 20063 21169: contig of 1107 bp in length
* 21170 21269: gap of unknown length
* 21270 22452: contig of 1183 bp in length
* 22453 24582: gap of unknown length
* 24583 24583: contig of 2030 bp in length
* 24584 24583: gap of unknown length
* 24583 26383: contig of 1701 bp in length
* 26384 26483: gap of unknown length
* 26484 27710: contig of 1227 bp in length
* 27711 27810: gap of unknown length
* 27811 29195: contig of 1385 bp in length
* 29196 29295: gap of unknown length
* 29296 30800: contig of 1505 bp in length
* 30801 30900: gap of unknown length
* 30901 32555: contig of 1655 bp in length
* 32556 34653: contig of 1998 bp in length
* 34654 37122: contig of 2369 bp in length
* 37123 37222: gap of unknown length
* 37223 39506: contig of 2284 bp in length
* 39507 39606: gap of unknown length
* 39607 41220: contig of 1614 bp in length
* 41221 42978: contig of 1658 bp in length
* 42979 43078: gap of unknown length
* 43079 45774: contig of 2695 bp in length
* 45775 45873: gap of unknown length
* 45874 55056: contig of 9183 bp in length
* 55057 55156: gap of unknown length
* 55157 67059: contig of 11903 bp in length
* 67060 67159: gap of unknown length
* 67160 80260: contig of 13101 bp in length
* 80261 80360: gap of unknown length
* 80361 96528: contig of 16168 bp in length
* 96529 96628: gap of unknown length
* 96629 114530: contig of 17901 bp in length
* 114531 114629: gap of unknown length
* 114630 133673: contig of 19044 bp in length
* 133674 133773: gap of unknown length
* 133774 153791: contig of 20018 bp in length
* 153792 153891: gap of unknown length
* 153892 176080: contig of 22188 bp in length
* 176081 207815: contig of 31636 bp in length.
*
FEATURES
      source
      1..207815
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="1"
      /clone="Rp11-243G22"
      misc_feature
      1..1381
      /note="assembly_name:Contig13"
      misc_feature
      1482..2602
      /note="assembly_name:Contig15"
      misc_feature
      2703..4013
      /note="assembly_name:Contig21"
      misc_feature
      4114..5331
      /note="assembly_name:Contig25"
      misc_feature
      5432..6622
      /note="assembly_name:Contig33"
      misc_feature
      6723..8166
      /note="assembly_name:Contig37"
      misc_feature
      8267..9674
      /note="assembly_name:Contig38"
      misc_feature
      9775..10943
      /note="assembly_name:Contig40"
      misc_feature
      11044..12198
      /note="assembly_name:Contig42"
      misc_feature
      12299..14193
      /note="assembly_name:Contig45"
      misc_feature
      14294..15717
      /note="assembly_name:Contig47"

```



```
Db 109814 GCTGTTGACCTTCTCATCAGCTCCATCAGAGGGCCAGGTCCTCGAAGGCTCTG 109873
QY 141 ggaacctgggtccagtgccctcttctccagccctgcggggt-99999tgcagcgca 199
Db 109874 GGGACCTTGGGTTCGATGGGCTTCTGCTCCAGAGCCCTGTGGGTGGGGGTGCAGCGCA 109933
QY 200 ggaagccggagatgtagctccctctacagtcagctccacccgagtgctgcccctccctccccc 259
Db 109934 NGAGCGACATGT-----GAACTGCACCCAGCCCTACCCCTCCCTCCCC 109978
QY 260 ggcccccaagacatccagaagccctccctccctcccccggggccaggggtcccccagaccccgactt 319
Db 109979 GACCCCAAGACATCAGAAAGCCACCGCGCCCGGGCCAGGCTCCAGACCCAGAGCTC 110038
QY 320 ctccagaacacctccctctgtagagacacagctctcgggaaggggtgcccacttcag 379
Db 110039 CAGGGATCCCAAGTCTCTGTATAGCCACAGCTTCGGGAAGGGGTGCCCTCTTCGAG 110098
QY 380 gtcccgcttccacacctagggagagagagacccagagatctcgagggccagg----- 432
Db 110099 CTCTGCTTCTCAAGTGGGAGAGAGAAACCCAGAGGCTCAAGGGGCCAGAGGTGAG 110158
QY 433 ----- 432
Db 110159 GCTCTAGGGCAGAGCAGAAAGGGGTAAATGGAAGGGCTTCAGAGGCCAAATCTTTC 110218
QY 433 -----aggctcc 438
Db 110219 AGAAGCAGACTACGGAGACAGGCTAAGAGATGACTAAGTCTACACATCTCTAGGTTC 110278
QY 439 cggcttcgagacccctcagcaggaatgttcgttatggagagtgccctttcattg 498
Db 110279 CGGTTTCAGAGACCCCATCAGCCAGGGATGTTCCGGTTATGGAGAGTGCTTTTGCCTTG 110338
QY 499 ccactgcacgggaacccgagccacctcgagccaccagagactgtgagctgtccctgact 558
Db 110339 CCCTGCACCGGAGCCGCGGATCCT-----CACAGACCCGGGCAACCCCAAGAAC 110389
QY 559 tctctagaggggaagagcttctccgtccctactcccaagacagagaccattctccgca 618
Db 110390 TCTTCTACAGGGGAAGGATGTTCCATCTCAGCTTCCAGACAGAGTGTCGCCCTCTGAA 110449
QY 619 aacggcagcccccaactgactcctcccccagaaactgtctgtccacaccccccactcccc 678
Db 110450 AAGCAGGGCCCTCATATGAGCCGCCCTG-----AACCCGGTCTCTAC 110491
QY 679 caagcagaaccttaagcctgaactgtctcagacagagtggtgcccacgaagccaggtcc 738
Db 110492 TCAGCTGAACCTCCAAGGTTGGAAGTGTCTCAGACAGAGAGTGCTCCCAAGAACTAGTTCT 110551
QY 739 gccccctacggcatacccccagagcccaggccctctggcacagagccccctccaccag 798
Db 110552 GCTCCT-----TCCTACACAGGGGACCCCTGCACCTACC 110584
QY 799 cactccttaggagaagtggtcttctccgtgcatccctcagcccaaggaagccaggtcc 858
Db 110585 TCCTTTTGGACAGAGTAGGTTTTCAGAGGATCCCTTGGGCCACGAATGCCACCTCC 110644
QY 859 cagg---gttggccagtcctccaggttagcagggagacgcccgtgactttctctggctc 915
Db 110645 CCGGGAAGTTGGAGCAGTCCCGAGGAGCAGAGCGAGCAGAC---ACCCACCTCTTCTCT 110701
QY 916 cctcggggccagggccagggcccaagggcccttggggaaacgggggactccctcaaggg 975
Db 110702 CTTGTCCCGGAGCAGCAGAGCAGAGGAGGACACTGGAGACCTCCAGAGTCTCTCACAG 110757
QY 976 ccccgctggagcctgacctcagccccggggcgctgctgctgcccctgctgagcaagcgc 1035
Db 110758 -----GTCCCAAGATGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110791
QY 1036 cccatgcagagtcctctgagagcctcttctgtccagtagccctattccaagatgttct 1095
```

```
Db 110792 -----TCCAGCCACACTCTGGAGCATCTTCTCCAGTATCCCTGCTCCAAACTGTTCT 110845
QY 1096 ggcagagtaaacagtaagagcctgcagccaagcg 1131
Db 110846 GGGAAAGTGAACAGATCGGAGCCTGAGCCAGGAG 110881

RESULT 8
BC008840 2703 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, hypothetical protein FLJ13710, clone MGC:14171
DEFINITION IMAGE:4120678, mRNA, complete cds.
ACCESSION BC008840
VERSION BC008840.1 GI:14250741
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc_mgc@nhgri.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgueon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 20 Row: C Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 10435805.
Location/Qualifiers
1. .2703
/organism="Homo sapiens"
/db_xref="LocusID:79875"
/db_xref="taxon:9606"
/clone="MGC:14171 IMAGE:4120678"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
373. .1737
/codon_start=1
/product="hypothetical protein FLJ13710"
/protein_id="AAH08840.1"
/db_xref="GI:14250742"

CDS
```

FEATURES  
source

QY	2244	gcagtcgcggcaggaatttgaggggggtgcctcctcagtgcccccggagcgcgtgtagaca	2303
Db	1035	TCGTGTGCCCGCCAGGTGTACGCCAACCCAGCCCTGACGGTGCACCCCTACCGTGCACGA	1094
QY	2304	ttctccccggcccaacatcacccagcttctgccagctgcgcctctgtggccattggggaagt	2363
Db	1095	CCTGGAGAAACCTGAGACCACCAGCACCCTGCCAACTCAAGATCGACGGAGTGGCAGAT	1154
QY	2364	tggctctcttggagccaagtctccgtgcgggtgagggccggggccagagagaagcggcaggt	2423
Db	1155	CCGGACCGACATGGACCTGCTCGGTGCCCTGCGGGCGTGGGACAGAGACCCTGATGT	1214
QY	2424	tgcgtgtgttgggaacaacagctgatgaagtgcgagcagcagagtgctgcgtcaggccccc	2483
Db	1215	GAAGTGTGTGACCAACATGGGGATGTGTGACGATGAGGAATGCAACATGAAGTC--	1272
QY	2484	acagccccccagcagagaggccttgatcacatggggccctgtactactgtcgtgtccacag	2543
Db	1273	-CGGCCGAATGACATTGAGAACTCGACATGGGACCTGTGCCAAGAGCTGTTCTCTCAC	1331
QY	2544	cgaactggagc	2553
Db	1332	CGAGTGGAGC	1341
RESULT	9		
LOCUS	AK023772	2964 bp	linear
DEFINITION	Human sapiens cDNA FLJ13710 fis, clone PLACE2000373, weakly similar to F-SPONDIN PRECURSOR.		
ACCESSION	AK023772		
VERSION	AK023772.1	GI:10435805	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Human sapiens placenta cDNA to mRNA, clone_lib:PLACE2 clone:PLACE2000373.		
ORGANISM	Human sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (sites)		
	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Naganari,K., Masuno,Y. and Oshima,A.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 2964)		
AUTHORS	Isogai,T. and Otsuki,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'-, 3'-end one pass sequencing and clone selection Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.		
FEATURES	Location/Qualifiers		
source	1. 2964		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="PLACE2000373"		
	/tissue_type="placenta"		
	/clone_lib="PLACE2"		
	/note="cloning vector: pME18SFL3"		
CDS	96..1712		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAB14673.1"		
	/db_xref="GI:10435806"		
	/translation="LFWYKRPNEISSTAGESFLAEGPTNEILDVYHIQHPNGVHH"		





QY	1418	gctgtggagctgtgtgggggtgatattctactctgcctctgttttcggggaaacctcactg	1477
Db	577	AGTGTCTGCAGTGTGGGGGTGATGGCTCATCTCTACTCCGGGTACAGGAACCTTTGATG	636
QY	1478	accgagggggcccctggcctgatcagaagactctgtggattccagcgggagccttcgccc	1537
Db	637	GCAATGATCTCAGCAGAGGCTACAAACCAAACTTCATCATTCGGCTGGGGCAACTAGTA	696
QY	1538	tccagattcccagctccgggocctagctccaaactacctggcaacttcgtgcccctggggccc	1597
Db	697	TTTCGATTTGAGGAAGCCCTGCCAGCAGAAATTTCTGGCGGTGAAGAGATCCGTGGCG	756
QY	1598	ggctccatcacaatgggaactcggctgtgggtcccccctggctcctacagcggcggcgggga	1657
Db	757	AGTACTATCTCAATGGGCACTGGACCATTAGAGCAGCCAGGCTCTTCGGTGGCCAGCA	816
QY	1658	ccgtcttttgatataaaccgctcctcccaggaggagggaagggagagctctgtcggctg	1717
Db	817	CAGTCTCAGTACGAACGGGGTGTGGAGGGGGACCTGGCTCTGAGCGGCTCCAGGCAC	876
QY	1718	aaggccccacaccagcctgtgatgtctatgatcttcaggagagaaaacccaggcg	1777
Db	877	GGGGCCCCACCTCAGAGCCCCGTGTCATCGAGCTCCTGAGTCAGGATCTCAACCCCGGAG	936
QY	1778	ttttttcaagtatgtcatctcttcaacctctccaatccttgagaacccccaccacccagagc	1837
Db	937	TACACTACGAGTACTACTGCCCGGAATGATCTCTGGCC----GAGGCTTCAGCTGGAGC	992
QY	1838	cccctgtccccagcttcagccggagattctgaggggtggagcccccaactgctccggcac	1897
Db	993	CATGGCTCTCTGGGAGACTGCAAGCGCTGAGTGTGTGGAGTCAACCAAGTCCGGCTGGTA	1052
QY	1898	cccgcgccagccggacccccagggcacccctccagcgtccaggtcgcgatcccccatgcccg	1957
Db	1053	TTCTGCACATTGACAACGAGGCCATTCCAGACCACATGTGCCAACACCAGCCTCGGCC	1112
QY	1958	cccgcgcccatccccagg----acacctcgggtctccagctcggtactggaaacagatg	2013
Db	1113	ACCCACCGAGCTTCATGCAACACTCAACTGTGTCGAAGACCAAGCGCTGGAAGGTAGGC	1172
QY	2014	ggacactctgcattgcagcgtccctgcgggaaagtgctctgcccgcacatttctcctgc	2073
Db	1173	CCATGAGCCCTGCTCGGTCTCCTCGGGGTGGAGTCCAGTCTCGTCTGCTACTGCTC	1232
QY	2074	atctccctgagtcgggagaggaactgtagt	2104
Db	1233	ATATCTTCAGATGGGACTGGTGCCAGGAAG	1263

```

RESULT 11
AX319854
LOCUS
DEFINITION Sequence 18 from Patent WO0183782.
ACCESSION AX319854
VERSION AX319854.1 GI:17901444
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and
Payne,V.
TITLE Novel proteases
JOURNAL Patent: WO 0183782-A 18 08-NOV-2001;
Sugen, Inc. (US)
FEATURES
source
1..2805
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
745 a 677 c 733 g 650 t
BASE COUNT
ORIGIN

```

Query Match 4.3%; Score 113.2; DB 6; Length 2805;  
 Best Local Similarity 50.0%; Pred. No. 1,7e-13;  
 Matches 317; Conservative 0; Mismatches 308; Indels 9; Gaps 1

QY	1159	cgggccctgcagtcgcagcctttaactccagaaattcatggccagctctatcagtg	1218
Db	1039	CGGGCTCAACAGTGTGCAGAAATATACAGCAAACTTCCGTGGATGTTCTACAGTGG	1098
QY	1219	gagcccttcaactgaagtcacagggtcccagcgcgtgtgaactgaactccgcccctggc	1278
Db	1099	AAACCCATACAAAAGTGAAGAGGAAGATCGATGCAAACTGTACTCAAGCCTGAGAAC	1158
QY	1279	ttccgcttctatgtccgtcacactgaaaagtcaggatgggaacctgtgtcacccctgga	1338
Db	1159	TTTGAATTTTITTTTGGCAATTCGCGCAAGTGAAGATGAATCCCTGTCCTCCCAAC	1218
QY	1339	gccctgacactctgtgtggctggagcgtgtgtgagcccgcgtgtgatgggatacctggc	1398
Db	1219	AAAAATGATGTTTGTATTGACGGGTTTGTGAAC TAGTGGGATGTGATCATGAAC TAGGC	1278
QY	1399	tctggcaggcgtccctgatggctgtgtggagctgtgtgggggtgatgattctacctgtgcctt	1458
Db	1279	TCTAAGCAGTTTCAGATGCTGTGGCGTTTGC AAAGGTGATAATCAACTTGCAAGTTT	1338
QY	1459	gttcggggaaacctcaactgaccgagggggccccctgggctatcagaagatcttgtggatt	1518
Db	1339	TATTAAGCCTGTACCTCAAC CAGATAAAGCAAAATGAATATTATCCGGTGTCTCATTT	1398
QY	1519	ccagogggagccttgcgctccagattgccagctccggctagctccaactacctggca	1578
Db	1399	CCAGTGGCGGCCGAAGCATCGAAATCCAGAGTGCAGGTTTCTCCAGATTACCTCGCA	1458
QY	1579	cttcgtggccctgggggcgggtcccaatcaatgggaactggcgtgtggatccccctggg	1638
Db	1459	GTTTCAACCTTCAGTCAAAAGTATTACCTACCGGGGGCTGGAGATCGACTGGCCTGGG	1518
QY	1639	tctcacagggccgggggaccgtcttttcgatataaacctctctcccagggaggggcaaa	1698
Db	1519	GAGTTCCTCCTTCGCTGGGACCACGTTTGAATACCAACGCTCTTTCAA-----CCGC	1569
QY	1699	gggagagactctgtggctggaagccccaccacccagctgtggatgtctatatgatcttt	1758
Db	1570	CGGNAACGTCTGTACGCGCCAGGGGCCACAAATGAGACGCTGGTCTTTGAAATTCGTATG	1629
QY	1759	caggagggaaaacccaggcgttttttatcagtatg	1792
Db	1630	CAAGCAAAAATCCAGGGATAGCTTTGGGAAGTATG	1663

RESULT	12
AX319852	
LOCUS	AX319852 3675 bp DNA linear PAT 14-DEC-2001
DEFINITION	Sequence 16 from Patent WO0183782.
ACCESSION	AX319852
VERSION	AX319852.1 GI:17901442
KEYWORDS	human.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (sites)
AUTHORS	Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and Payne,V.
TITLE	Novel proteases
JOURNAL	Patent: WO 0183782-A 16 08-NOV-2001;
FEATURES	Sugen, Inc. (US) Location/Qualifiers 1..3675 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	874 a 1009 c 1019 g 773 t



QY	1718	aaggccccaccaccagcctgttgatgtctctatatgtctcttcaggaggagaaacccagcg	1777		
Db	9464	CGGTCCTCCATCTCGGAGAGCCTCTTTCATAGTTATGTTGGTCAGAGAGAACATCAGCC	9523		
QY	1778	ttttttatcagttatgcat	1796		
Db	9524	TCGACTACGAGTACAGAT	9542		
RESULT 15	AC009213	182525 bp	DNA	linear	INV 17-FEB-2001
LOCUS	AC009213/c				
DEFINITION	Drosophila melanogaster, chromosome 3R, region 98D-98D, BAC clone				
ACCESSION	EACR09F18				
VERSION	AC009213				
KEYWORDS	AC009213.5	GI:12957654			
SOURCE	HTG.				
ORGANISM	fruit fly.				
REFERENCE	Drosophila melanogaster				
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
	1 (bases 1 to 182525)				
	Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Ananides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Busan,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Hock,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacלב,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.				
TITLE	Sequencing of Drosophila chromosome 3R, region 98D-98D				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 182525)				
AUTHORS	Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chaviez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacלב,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA				
COMMENT	On Feb 17, 2001 this sequence version replaced gi:6957913.				
	Sequence submitted by:				
	Berkeley Drosophila Genome Project				
	Lawrence Berkeley National Laboratory, MS 64-121				
	Berkeley, CA 94720				
	This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site ( <a href="http://www.fruitfly.org/sequence/">http://www.fruitfly.org/sequence/</a> ) or send email to <a href="mailto:bdg@fruitfly.berkeley.edu">bdg@fruitfly.berkeley.edu</a> .				
FEATURES	Location/Qualifiers				
source	1. 182525				
	/organism="Drosophila melanogaster"				
	/strain="y: cn bw sp"				
	/db_xref="taxon:7227"				
	/chromosome="3R"				
	/map="98D-98D"				
	/clone="EACR09F18 (9812)"				
	/clone.lib="RPCI-98 (Roswell Park Cancer Institute				
	Drosophila melanogaster BAC library, partial EcoRI in				

BASE COUNT 49013 a 41841 c 41897 g 49774 t  
ORIGIN pBACE3.6)"

Query Match 3.7%; Score 97; DB 3; Length 182525;  
Best Local Similarity 47.6%; Pred. No. 2e-10;  
Matches 323; Conservative 0; Mismatches 350; Indels 6; Gaps 1;

QY 1118 cctgagcaagcgcctcccccctgagcagcagccccccggccctgcaagtgcgag 1177  
|||||  
Db 70744 CTTGCAACACCCAAAGACTTCCTGGAGAGAAATCCGACTTCGGGGCGAGAGTGTCTCC 70685  
QY 1178 cctttaactcccaggaattcatgggccaagtgtatcagtggagcccttcaactgaagtcc 1237  
|||||  
Db 70684 GCTTCGACAGGAGCAGTTTGTAGTCTTCTACGAGTGGGTGCGGTACACGAATGCAC 70625  
QY 1238 agggctccagcgtgctgaactgaactgcggcccccgtggttcgcttctatgtccgtc 1297  
|||||  
Db 70624 -----CCAACTCCCTCGAGCTGAATGTATGCCAAAGGGCGAGCGGTCTACTACCGTTC 70571  
QY 1298 acactgaaaggtccaggaatgggacccctgtcagcgtgagccccctgacatctgtgtgg 1357  
|||||  
Db 70570 AGCGGAGAGGTGGTGAGCGGACCCGCTGCACGACAAAGACCTCGACGTGTGCGTAA 70511  
QY 1358 ctggagcgtgtctgagccccggcgtgtgatggatcccttggctctggcaggggcctgtatg 1417  
|||||  
Db 70510 ACGGGGAGTGCATCGCGTGGGTGGCTGCACATGCTGGGTAGCGATGCCAAAGGAGGACA 70451  
QY 1418 gctgtgagctgtggggtgatgattctacctgtgcctgttttcggggaaacctcactg 1477  
|||||  
Db 70450 AGTCCGAAAGTGTGTGTGTGATGGCAGACCTGCNAGACCAATTCGCAACACCATCACA 70391  
QY 1478 accgagggggccccctggcctatcagaagattgttggtattccagogggagccttgcggc 1537  
|||||  
Db 70390 CGAAGGATCTGCTCTGGATACAATGACTGTGTCTCTCCAGAGGGAGCCACTAACA 70331  
QY 1538 tcagattccagcgtccggcctagctccaactaccctgaccttcgtggccctggggcc 1597  
|||||  
Db 70330 TTGCTATCGAGGAGACTGTGCCCTCCAGCAATTACCTGGCTCCCGGAACACAGTGGAC 70271  
QY 1598 ggtccatcatcaatgggaactgggctgtggatccccctgggtccctacagggccgcgggga 1657  
|||||  
Db 70270 ACTACTACCTGAACGGCGATTGGCGCATCGACTTCACAGCCCATGTTCTTCGGGAAC 70211  
QY 1658 cegtcttcgatataaccgtccctccagggaggggcaaggagagctgtcggtg 1717  
|||||  
Db 70210 CGTGGTGGAACTACCAACGAAGGCCAATGGGCTTCGCCGCTCCCGACCAACTGACTGCA 70151  
QY 1718 aagccccccaccagcctgtggatgtctatatgatctttcaggaggaaacccaggcg 1777  
|||||  
Db 70150 GCGTCCCATCTCGGAGAGCCCTTTCATAGTTATGTTGCTGCAGGAGAGAACAATCAGCC 70091  
QY 1778 tttttatcagtatgcat 1796  
|||||  
Db 70090 TCGACTACGAGTACGAT 70072

Search completed: July 23, 2002, 17:35:57  
Job time: 16247 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 18:27:20 ; Search time 292.12 Seconds  
(without alignments)  
15481.147 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2834  
Sequence: 1 atgaggaactggactggcag.....ccctagcctttccagcatag 2634

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1259	47.8	2680	22 AAC91178	ADAM gene #2. Hom
2	377	14.3	1020	22 AAD16773	Human novel protei
3	377	14.3	1998	22 AAD16758	Human novel protei
4	54	2.1	320	21 AAC16560	Human secreted pro
5	23	0.9	704	22 AAH93398	Human protein enco
6	23	0.9	814	22 AAH08271	Human CDNA clone (
7	23	0.9	2964	22 AAH16636	Human CDNA sequenc
8	20	0.8	893	23 AAS82152	DNA encoding novel
9	20	0.8	1827	23 AAS66958	DNA encoding novel

10	20	0.8	1927	22 AAH76376	Murine adipocytes-
11	20	0.8	1949	22 AAC84391	Mouse A236 polypep
12	20	0.8	1949	22 AAC84416	Mouse A236 variant
13	20	0.8	1949	22 AAC84417	Mouse A236 variant
14	20	0.8	1949	22 AAC84418	Mouse A236 variant
15	20	0.8	1995	22 AAF63462	Oligonucleotide #2
16	20	0.8	2013	23 AAS83442	DNA encoding novel
17	20	0.8	2153	23 AAS78212	DNA encoding novel
18	20	0.8	17069	22 AAK60906	Human immune/haema
19	20	0.8	17069	22 AAK83892	Human immune/haema
20	20	0.8	32206	22 AAK89374	Human digestive sy
21	19	0.7	308	22 AAS28913	Human immunoglobul
22	19	0.7	410	18 AAT51050	Glycoprotein gp105
23	19	0.7	410	21 AAH70394	Partial gp105 prot
24	19	0.7	609	22 AAL22948	Human breast cance
25	19	0.7	647	22 AAH06718	Human CDNA clone (
26	19	0.7	669	22 AAH16868	Human digestive sy
27	19	0.7	671	22 AAK88735	Human liver associ
28	19	0.7	671	22 AAS31769	Human CDNA encodin
29	19	0.7	724	22 AAS26853	Human reproductive
30	19	0.7	1790	22 AAL05360	Nucleotide sequenc
31	19	0.7	2055	19 AAV68524	DNA encoding a hum
32	19	0.7	2055	21 AA260905	Human notch ligand
33	19	0.7	2058	21 AA291643	Human CDNA sequenc
34	19	0.7	2159	22 AAS21287	A human delta-rel
35	19	0.7	2184	21 AA260918	Human CDNA sequenc
36	19	0.7	2437	22 AAH18557	Human CDNA SEQ ID
37	19	0.7	2442	22 ABA06484	Human CDNA sequenc
38	19	0.7	2508	22 AAH14855	Human CDNA sequenc
39	19	0.7	2615	22 AAH18513	Human CDNA sequenc
40	19	0.7	2670	23 ABL04247	Drosophila melanog
41	19	0.7	2800	19 AAV68523	Nucleotide sequenc
42	19	0.7	2897	24 AAD26454	Human kinase PKIN-
43	19	0.7	3339	20 AAX16297	Human delta-2 prot
44	19	0.7	4426	23 ABL17454	Drosophila melanog
45	19	0.7	4878	23 ABL04246	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAC91178  
ID AAC91178 standard; DNA; 2680 BP.  
XX  
AC AAC91178;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE ADAM gene #2.  
XX  
KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
KW ovarian; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200073323-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 25-MAY-2000; 2000WO-US14308.  
XX  
PR 27-MAY-1999; 99US-0136388.  
PR 09-JUL-1999; 99US-0142930.  
PR 28-JAN-2000; 2000US-0176717.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;  
XX  
DR WPI; 2001-016507/02.  
XX  
PT Seven nucleic acid molecules encoding ADAM polypeptides containing a

PT disintegrin and metalloprotease domain, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
XX Claim 1; Page 266; 287pp; English.  
XX  
CC The present invention relates to seven members of the ADAMs (proteins  
CC which contain A Disintegrin And Metalloprotease domain) protein family.  
CC The ADAMs proteins and DNA may be used to treat disease, as a food  
CC additive or preservative, for chromosome identification, as probes  
CC for diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the  
CC gene therapy of breast and ovarian cancer.  
XX  
XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

Query Match 47.8%; Score 1259; DB 22; Length 2680;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1301 ctgaaaggtccaggatggagacctgtgcagcctggagccctgacatctgttggtcg 1360  
Db 3 ctgaaaggtccaggatggagacctgtgcagcctggagccctgacatctgttggtcg 62  
QY 1361 gacgtctgtgagcccgctgtgatggatccttgctcagcagcctcctgatgctt 1420  
Db 63 gacgtctgtgagcccgctgtgatggatccttgctcagcagcctcctgatgctt 122  
QY 1421 gtggagctgtgggggtgatgttacctgtgcctctgttcgggaaacctcaactgacc 1480  
Db 123 gtggagctgtgggggtgatgttacctgtgcctctgttcgggaaacctcaactgacc 182  
QY 1481 gagggggccccctgggctatcaagaatctgtgattcagcagcagcctgcgctcc 1540  
Db 183 gagggggccccctgggctatcaagaatctgtgattcagcagcagcctgcgctcc 242  
QY 1541 agatgcccagctccgctcagtcacactacctggaactctgctgcccctggggccggt 1600  
Db 243 agatgcccagctccgctcagtcacactacctggaactctgctgcccctggggccggt 302  
QY 1601 ccatcatcaatggaaactggctgtggatccccctggctcctacagggccgggagccg 1660  
Db 303 ccatcatcaatggaaactggctgtggatccccctggctcctacagggccgggagccg 362  
QY 1661 tcttcgataaacctcctccagggagaggcgaaaggagagagctctgctggctgaag 1720  
Db 363 tcttcgataaacctcctccagggagaggcgaaaggagagagctctgctggctgaag 422  
QY 1721 gcccccacacccagcctgtggtatgttatatgatctttcaggaggaacccagcgctt 1780  
Db 423 gcccccacacccagcctgtggtatgttatatgatctttcaggaggaacccagcgctt 482  
QY 1781 ttatcagtagtgcattctctccctccctccatccttgagacccacccagagccccc 1840  
Db 483 ttatcagtagtgcattctctccctccctccatccttgagacccacccagagccccc 542  
QY 1841 ctgtcccccagctccagcggagattctgaggggtggagcccccactgtccggacccc 1900  
Db 543 ctgtcccccagctccagcggagattctgaggggtggagcccccactgtccggacccc 602  
QY 1901 gccacccgggacccagcagcctccagctcaggtcgagatcccccagatcccgccc 1960  
Db 603 gccacccgggacccagcagcctccagctcaggtcgagatcccccagatcccgccc 662  
QY 1961 cgcccattccagagacacccctgggtctccagctcgtactggaaacagtgaggacact 2020  
Db 663 cgcccattccagagacacccctgggtctccagctcgtactggaaacagtgaggacact 722  
QY 2021 ctgcattcagcgtcctcgggaaagggtgtctgggcccacatttctctcattctccc 2080  
Db 723 ctgcattcagcgtcctcgggaaagggtgtctgggcccacatttctctcattctccc 782

QY 2081 gtgagtcgggagagaaactgatgaacacagctgtgcgcgggtgcccagggcccccagct 2140  
Db 783 gtgagtcgggagagaaactgatgaacacagctgtgcgcgggtgcccagggcccccagct 842  
QY 2141 cccctgaacccctgcacagcgcaccccatgccccccatactgggagggctggcagtggaat 2200  
Db 843 cccctgaacccctgcacagcgcaccccatgccccccatactgggagggctggcagtggaat 902  
QY 2201 cctgcagcgcctcctctgtgccccggcaccacagcaccgacagctcagtgccggcaggaat 2260  
Db 903 cctgcagcgcctcctctgtgccccggcaccacagcaccgacagctcagtgccggcaggaat 962  
QY 2261 ttgggggggggtggctcctcctggtgccccgggagcgtgtggacatctccccgggcccaca 2320  
Db 963 ttgggggggggtggctcctcctggtgccccgggagcgtgtggacatctccccgggcccaca 1022  
QY 2321 tcacccagctcttgccagctgcgcctctgtgcccattgggaagtgtgctctcttgagacc 2380  
Db 1023 tcacccagctcttgccagctgcgcctctgtgcccattgggaagtgtgctctcttgagacc 1082  
QY 2381 agtgcctcgtcggtgctgcggcgccgggcccagaagaacccgaggttcgtgtgttgggaaca 2440  
Db 1083 agtgcctcgtcggtgctgcggcgccgggcccagaagaacccgaggttcgtgtgttgggaaca 1142  
QY 2441 acggtgatgaagttagcagcagcagagtgctgcagggccccccacagccccccagcagag 2500  
Db 1143 acggtgatgaagttagcagcagcagagtgctgcagggccccccacagccccccagcagag 1202  
QY 2501 aggcctgtgacatggggccctgtactactgctgttccacagcagctggagctcccaag 2559  
Db 1203 aggcctgtgacatggggccctgtactactgctgttccacagcagctggagctcccaag 1261

RESULT 2  
AAD16773  
ID AAD16773 standard; cDNA; 1020 BP.  
XX AC AAD16773;  
XX DT 22-NOV-2001 (first entry)  
XX DE Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.  
XX KW Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;  
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
KW anti-sense therapy; endocrine disorder; leukaemia; ss.

XX OS Homo sapiens.  
XX PH Key Location/Qualifiers  
XX CDS 65..847  
XX FT /\*tag= a  
XX FT /product= "Human novel protein"  
XX FT /transl\_except= (pos:767..769, aa:Xaa)  
XX FT /note= "CDS does not include start codon; Xaa is an  
XX FT unknown amino acid"  
XX FT /partial  
XX PN WO200155202-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01325.  
XX PP 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.









RESULT 4  
AAC16560 standard; cDNA; 320 BP.  
XX AC AAC16560;  
XX DT 06-OCT-2000 (first entry)  
XX XX Human secreted protein 5' EST, SEQ ID NO: 20635.  
XX DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX KW gene therapy; chromosome mapping; ss.  
XX KW Homo sapiens.  
XX OS EF1033401-A2.  
XX PN 06-SEP-2000.  
XX PD 21-FEB-2000; 2000EP-0200610.  
XX PF 26-FEB-1999; 99US-0122487.  
XX PR (GEST) GENSET.  
XX PA Dumas Milne Edwards J, Duclert A, Giordano J;  
XX PI WPI; 2000-500381/45.  
XX DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX XX  
XX PS Claim 1; SEQ ID 20635; 71pp + CD-ROM; English.  
XX CC The present sequence is one of a large number of 5' ESTs derived from  
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX CC identified within the present sequence. The 5' ESTs were prepared from  
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX CC derived from the 5' ends of mRNAs and even in those cases where longer  
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX CC They are used to obtain upstream regulatory sequences and to design  
XX CC expression and secretion vectors.  
XX SQ Sequence 320 BP; 50 A; 97 C; 117 G; 53 T; 3 other;  
  
Query Match 2.1%; Score 54; DB 21; Length 320;  
Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 atgagaaactgactggcagccctggtgtatctgtgctgtctgtctgtctc 54  
|||||  
Db 267 atgagaaactgactggcagccctggtgtatctgtgctgtctgtctc 320  
|||||  
  
RESULT 5  
AAH99398  
ID AAH99398 standard; cDNA; 704 BP.  
XX AC AAH99398;  
XX DT 16-OCT-2001 (first entry)  
XX DE Human protein encoding cDNA sequence SEQ ID NO:233.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnery; antileukic; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX Homo sapiens.  
XX OS WO200153455-A2.  
XX PN 26-JUL-2001.  
XX PD 22-DEC-2000; 2000WO-US35017.  
XX PF 23-DEC-1999; 99US-0471275.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-0552317.  
XX XX (HYSE-) HYSEQ INC.  
XX XX Tang YT, Liu C, Drmanac RT;  
XX PI WPI; 2001-457603/49.  
XX DR P-PSDB; AAM25457.  
XX XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX PT  
XX Claim 1; Page 401-402; 1217pp; English.  
XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX CC AAM25963. The proteins can have activities based on the tissues and  
XX CC cells they are expressed in, such as: antinflammatory; antirheumatic;  
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
XX CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
XX CC antileukic; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine  
XX CC production. The proteins and polynucleotides are useful for screening for  
XX CC agonists or antagonists of a protein and for the treatment and diagnosis  
XX CC of disorders associated with the activity of a protein e.g. inflammation,  
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
XX CC neurological disorders.  
XX SQ Sequence 704 BP; 184 A; 158 C; 195 G; 167 T; 0 other;

Query Match 0.9%; Score 23; DB 22; Length 704;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1600 tccatcatcaatgggaactgggc 1622  
|||||  
Db 445 tccatcatcaatgggaactgggc 467

RESULT 6  
AAH08271  
ID RAH08271 standard; cDNA; 814 BP.  
XX  
AC AAH08271;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:5106.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
XX  
PR 27-AUG-1999; 99JP-0300253.  
XX  
PR 11-JAN-2000; 2000JP-0118776.  
XX  
PR 02-MAY-2000; 2000JP-0183767.  
XX  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 1; SEQ ID 5106; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 814 BP; 205 A; 235 C; 235 G; 136 T; 3 other;

Query Match 0.9%; Score 23; DB 22; Length 814;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1600 tccatcatcaatgggaactgggc 1622

Db 36 tccatcatcaatgggaactgggc 58  
|||||  
RESULT 7  
AAH16636  
ID AAH16636 standard; cDNA; 2964 BP.  
XX  
AC AAH16636;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:15752.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
XX  
PR 27-AUG-1999; 99JP-0300253.  
XX  
PR 11-JAN-2000; 2000JP-0118776.  
XX  
PR 02-MAY-2000; 2000JP-0183767.  
XX  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 15752; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

Query Match 0.9%; Score 23; DB 22; Length 2964;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1500 tccatcatcaatgggaactgggc 1622  
|||||  
Db 36 tccatcatcaatgggaactgggc 58

RESULT 8  
AAS82152/c  
ID AAS82152 standard; cDNA; 693 BP.  
AC AAS82152;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #17956.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG17965.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 17956; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 693 BP; 159 A; 188 C; 204 G; 142 T; 0 other;

Query Match 0.8%; Score 20; DB 23; Length 693;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2229 ccagcacgccagctgcagt 2248  
|||||  
Db 75 CCAGCACGCCAGCTGCAGT 56

RESULT 9  
AAS66958  
ID AAS66958 standard; cDNA; 1827 BP.  
XX  
AC AAS66958;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #2762.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG02771.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 2762; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1827 BP; 374 A; 528 C; 512 G; 413 T; 0 other;

Query Match 0.8%; Score 20; DB 23; Length 1827;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2229 ccagcacgccagctgcagt 2248

Db 762 ccagcaccgcccagctgcagt 781  
|||||  
RESULT 10  
AAH76376  
ID AAH76376 standard; DNA; 1927 BP.  
XX  
AC AAH76376;  
XX  
DT 30-NOV-2001 (first entry)  
XX  
DE Murine adipocytes-derived protein encoding DNA.  
XX  
KW Adipocyte; murine; differentiation; obesity; hyperlipemia; diabetes;  
KW atherosclerosis; ds.  
XX  
OS Mus musculus.  
XX  
PH Key Location/Qualifiers  
FT CDS 289..1410  
FT /\*tag= a  
XX  
PN WC200166720-A1.  
XX  
PD 13-SEP-2001.  
XX  
PF 09-MAR-2001; 2001WO-JP01863.  
XX  
PR 10-MAR-2000; 2000JP-0072502.  
XX  
PA (KITA/) KITAMURA T.  
PA (TSUR/) TSURUGA H.  
XX  
PI Kitamura T, Tsuruga H;  
XX  
DR WPI; 2001-565585/63.  
DR P-PSDB; AAB85862.  
XX  
PT Genes associated with adipocyte differentiation for screening  
PT adipocyte-related disorders such as diabetes and hyperlipemia  
XX  
XX Claim 1; Page 68-72; 112pp; Japanese.  
XX  
CC The invention relates to genes derived from murine adipocytes and  
CC proteins encoded by these genes. The proteins are associated with  
CC adipocyte differentiation and can be expressed by standard recombinant  
CC methodology. The genes, proteins and specific antibodies are useful for  
CC the identification of drugs for treatment and prevention of adipocyte-  
CC related disorders such as obesity, hyperlipemia, diabetes and  
CC atherosclerosis. The present sequence represents a DNA encoding a  
CC protein derived from murine adipocytes.  
XX  
SQ Sequence 1927 BP; 509 A; 536 C; 517 G; 365 T; 0 other;  
  
Query Match 0.8%; Score 20; DB 22; Length 1927;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2276 cctcgggtgccccggagcgc 2295  
|||||  
Db 103 cctcgggtgccccggagcgc 122  
  
RESULT 11  
AAC84391  
ID AAC84391 standard; cDNA; 1949 BP.  
XX  
AC AAC84391;  
XX  
DT 02-APR-2001 (first entry)  
XX

DE Mouse A236 polypeptide encoding cDNA.  
XX  
KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;  
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;  
KW antiasthmatic; neuroprotective, cytostatic; cardiant; hepatotropic;  
KW antiinflammatory; antidiabetic; antifertility; antipyretic; vasotropic;  
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;  
KW ophthalmological; antiskilling; antiulcer; vulnary; ss.  
XX  
OS Mus sp.  
XX  
XX Key Location/Qualifiers  
FT CDS 304..1425  
FT /\*tag= a  
FT /product= "mouse A236"  
XX  
PN WO200069895-A2.  
XX  
PD 23-NOV-2000.  
XX  
PF 15-MAY-2000; 2000WO-US13361.  
XX  
PR 14-MAY-1999; 99US-0312359.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Pan Y, Leiby KR;  
XX  
DR WPI; 2001-024999/03.  
DR P-PSDB; AAB48126.  
XX  
PT Novel nucleic acids encoding secreted or transmembrane proteins, useful  
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of  
PT the lung, liver, kidney or pancreas -  
XX  
XX Claim 2; Fig 24A-D; 209pp; English.  
XX  
CC The invention provides human and mouse nucleic acids designated TANGO  
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane  
CC proteins. The polypeptides, nucleic acids and their modulators may be  
CC useful for treating or modulating cholesterol uptake, blood coagulation,  
CC to modulate cell proliferation, morphogenesis and fate specification,  
CC tissue repair and renewal to treat cancer and promote wound healing,  
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan  
CC syndrome, protein S deficiency, modulate allergic or inflammatory  
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and  
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,  
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, osteoporosis,  
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain  
CC herniations, meningitis, ischemic brain or heart disease, infarction,  
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart  
CC disease, pulmonary heart disease, rheumatic fever, congenital heart  
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,  
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,  
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's  
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,  
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune  
CC Albright syndrome, infertility, uterine disorders, viral disease. The  
CC present sequence represents the mouse A236 cDNA.  
XX  
SQ Sequence 1949 BP; 508 A; 550 C; 527 G; 364 T; 0 other;  
  
Query Match 0.8%; Score 20; DB 22; Length 1949;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2276 cctcgggtgccccggagcgc 2295  
|||||  
Db 118 cctcgggtgccccggagcgc 137  
  
RESULT 12

```

AAC84416
ID AAC84416 standard; cDNA; 1949 BP.
XX
AC AAC84416;
XX
DT 02-APR-2001 (first entry)
XX
DE Mouse A236 variant 1 cDNA.
XX
KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;
KW antinflammatory; antidiabetic; antifertility; antipyrretic; vasotropic;
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
KW ophthalmological; antiskilling; antiulcer; vulnerary; variant; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 304..1425
FT FT /*tag= a
FT FT /product= "mouse A236 variant 1"
FT FT 366
FT FT /*tag= b
FT FT /*note= "wild-type G at this position is replaced with C"
XX
PN WO200069885-A2.
XX
PD 23-NOV-2000.
XX
PF 15-MAY-2000; 2000WO-US13361.
XX
PR 14-MAY-1999; 99US-0312359.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Leiby KR;
XX
DR WPI; 2001-024999/03.
XX
DR P-PSDB; AAB48148.
XX
PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT the lung, liver, kidney or pancreas -
XX
PS Claim 2; Page -: 209pp; English.
XX
CC The invention provides human and mouse nucleic acids designated TANGO
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC proteins. The polypeptides, nucleic acids and their modulators may be
CC useful for treating or modulating cholesterol uptake, blood coagulation,
CC to modulate cell proliferation, morphogenesis and fate specification,
CC tissue repair and renewal, to treat cancer and promote wound healing,
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC syndrome, acid secretion, tropic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, pulmonary heart disease, rheumatic fever, congenital heart
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a mouse A236 variant cDNA.
CC Note: the present variant sequence has been constructed using the
CC information provided in the specification.
XX
SQ Sequence 1949 BP; 508 A; 551 C; 526 G; 364 T; 0 other;

```

```

Query Match 0.88; Score 20; DB 22; Length 1949;
Best Local Similarity 100.0%; Pred. NO. 51;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2276 cctcggtgccccggagcgc 2295
DB 118 cctcggtgccccggagcgc 137
RESULT 13
AAC84417
ID AAC84417 standard; cDNA; 1949 BP.
XX
AC AAC84417;
XX
DT 02-APR-2001 (first entry)
XX
DE Mouse A236 variant 2 cDNA.
XX
KW TANGO 204; TANGO 206; TANGO 209; A236; secreted protein; human; mouse;
KW transmembrane protein; antianemic; cerebroprotective; arteriosclerosis;
KW antiasthmatic; neuroprotective; cytosolic; cardiant; hepatotropic;
KW antinflammatory; antidiabetic; antifertility; antipyrretic; vasotropic;
KW antirheumatic; nephrotropic; hemostatic; antilipemic; osteopathic;
KW ophthalmological; antiskilling; antiulcer; vulnerary; variant; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 304..1425
FT FT /*tag= a
FT FT /product= "mouse A236 variant 2"
FT FT 384
FT FT /*tag= b
FT FT /*note= "wild-type G at this position is replaced with C"
XX
PN WO200069885-A2.
XX
PD 23-NOV-2000.
XX
PF 15-MAY-2000; 2000WO-US13361.
XX
PR 14-MAY-1999; 99US-0312359.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Leiby KR;
XX
DR WPI; 2001-024999/03.
XX
DR P-PSDB; AAB48149.
XX
PT Novel nucleic acids encoding secreted or transmembrane proteins, useful
PT for treating, e.g. cancer, hemophilia, anemia, ischemia or diseases of
PT the lung, liver, kidney or pancreas -
XX
PS Claim 2; Page -: 209pp; English.
XX
CC The invention provides human and mouse nucleic acids designated TANGO
CC 204, TANGO 206, TANGO 209 and A236 encoding secreted or transmembrane
CC proteins. The polypeptides, nucleic acids and their modulators may be
CC useful for treating or modulating cholesterol uptake, blood coagulation,
CC to modulate cell proliferation, morphogenesis and fate specification,
CC tissue repair and renewal, to treat cancer and promote wound healing,
CC modulate angiogenesis, treat hypercholesterolemia, hemophilia, Marfan
CC syndrome, protein S deficiency, modulate allergic or inflammatory
CC response, acid secretion, tropic effects on gastrointestinal mucosa, and
CC promote ulcer healing, treat bone cancer, achondroplasia, myeloma,
CC leukemia, anemia, thalassemia, cerebral edema, hydrocephalus, brain
CC herniations, meningitis, ischemic brain or heart disease, infarction,
CC intracranial hemorrhage, pancreatitis, diabetes, angina, hypotensive heart
CC disease, pulmonary heart disease, rheumatic fever, congenital heart
CC disease, myocardial disease, atherosclerosis, hypertension, jaundice,
CC hepatic failure, cirrhosis, glomerulonephritis, Goodpasture's syndrome,
CC sickle cell disease, renal failure, ischemic bowel disease, Crohn's
CC disease, hernias, hypoadrenalism, hyperadrenalism, Cushing's syndrome,
CC neoplasia, pulmonary disorders, asthma, ovarian disorders, McCune
CC Albright syndrome, infertility, uterine disorders, viral disease. The
CC present sequence represents a mouse A236 variant cDNA.
CC Note: the present variant sequence has been constructed using the
CC information provided in the specification.
XX
SQ Sequence 1949 BP; 508 A; 551 C; 526 G; 364 T; 0 other;

```





CC hepatitis C virus (HCV) proteins. The method can be used to identify  
CC compounds which can be used for the prevention and treatment of liver  
CC disorders associated with HCV infection, such as liver cancer. The  
CC present sequence was used in the present invention.

XX  
SQ Sequence 1995 BP; 434 A; 594 C; 635 G; 332 T; 0 other;

Query Match 0.8%; Score 20; DB 22; Length 1995;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2229 ccagcacccgcagctgcagt 2248  
|||||  
Db 1357 CCAGCACCCGCAGCTGCAGT 1338

Search completed: July 23, 2002, 20:02:27  
Job time: 5707 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 20:02:31 ; Search time 59.24 Seconds

(without alignments)  
1644.357 Million cell updates/sec

Title: US-10-041-770-2

Perfect score: 4895

Sequence: 1 MENWTGRPWLYLLLSLPQ.....PPAISILGNHAQTSAPFA 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1554	31.7	491	22 AAB50936	ADAM protein #2.
2	858	17.5	260	22 AAE09713	Human gene 9 encod
3	858	17.5	372	22 AAE09698	Human gene 9 encod
4	817	16.7	150	22 AAB50947	ADAM gene #2 pepti
5	740.5	15.1	1690	22 AAB86949	Human metalloprote
6	735.5	15.0	1686	22 AAE00934	Human 27875 ADAM-T
7	735.5	15.0	1686	22 AAB74944	Human ADAM type me
8	734.5	15.0	1686	22 AAE00913	Human 27875 ADAM-T
9	722	14.7	634	22 AAU01291	Human Thrombospond
10	719	14.7	1103	22 AAB74945	Human ADAM type me
11	711.5	14.5	2858	22 ABB71150	Drosophila melanog

12	711.5	14.5	3060	22 ABB58064	Drosophila melanog
13	709	14.5	1072	22 AAB72300	Human ADAMTS-10 al
14	695	14.2	1081	22 AAB72288	Human ADAMTS-10 am
15	689	14.1	1284	21 ABA1379	Human ORFX ORF1143
16	661.5	13.5	538	22 AAB94727	Human protein sequ
17	626	12.8	523	22 AAO01292	Human Thrombospond
18	618	12.6	951	21 AAB21253	Human metalloprote
19	618	12.6	2150	21 AAY53698	Amino acid sequenc
20	618	12.6	2165	22 AAB90617	Human ADAMTS-9 alt
21	583	11.9	1934	22 AAB72301	Human ADAMTS-9 am
22	577.5	11.8	1882	22 AAB72286	Novel human diagno
23	573.5	11.7	634	22 AAG01904	Human secreted pro
24	556	11.4	1745	22 AAB90551	Human CDNA SEQ ID
25	556	11.4	1766	22 ABB10246	Human metalloprote
26	553	11.3	950	22 AAG62299	Novel human diagno
27	545	11.1	1328	22 AAG22373	Human TANGO 224 (f
28	542	11.1	874	21 AAB01431	Human metalloprote
29	540	11.0	958	21 AAB21255	Murine ADAMTS-10 a
30	537	11.0	450	22 AAB72289	Human N-proteinase
31	536	10.9	1211	19 AA47028	Human signal pepti
32	531.5	10.9	643	21 AAY87347	Bovine metalloprot
33	531	10.8	1203	22 AAB50004	Human ADAMTS-R1 am
34	530	10.8	525	22 AAB72290	Human ADAMTS-R1 am
35	528	10.8	525	20 AAY41729	Human PRO1071 (UNQ
36	528	10.8	525	21 ABA44385	Human PRO polypept
37	528	10.8	525	22 AAG29241	Novel human protei
38	524	10.7	1617	22 AAE07870	Human metalloprote
39	523.5	10.7	959	22 AAB86947	Novel human protei
40	522	10.7	771	22 AAE07869	Novel human protei
41	521.5	10.7	650	22 AAE07867	Human protease-rel
42	521.5	10.7	757	22 AAE03583	Human protease-rel
43	521.5	10.7	908	22 AAE03572	Human novel KIAA12
44	517	10.6	1691	22 AAE10609	Bovine N-proteinase
45	516.5	10.6	1206	19 AA47030	

ALIGNMENTS

RESULT 1

AAB50936  
ID AAB50936 standard; protein; 491 AA.

AC AAB50936;

XX 20-MAR-2001 (first entry)

XX ADAM protein #2.

XX ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
KW ovarian.

XX Homo sapiens.

XX WO200073233-A2.

XX 07-DEC-2000.

XX 25-MAY-2000; 2000WO-US14308.

XX 27-MAY-1999; 99US-0136388.

XX 09-JUL-1999; 99US-0142930.

XX 28-JAN-2000; 2000US-0178717.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;

XX WPI; 2001-016507/02.

XX Seven nucleic acid molecules encoding ADAM polypeptides containing a  
PT disintegrin and metalloprotease domain, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular

disorders and neurological diseases -  
Claim 11; Page 274-276; 287pp; English.

The present invention relates to seven members of the ADAMS (proteins which contain A Disintegrin And Metalloprotease domain) protein family. The ADAMS proteins and DNA may be used to treat disease, as a food additive or preservative, for chromosome identification, as probes for diagnosing a disorder related to the female reproductive system, particularly breast and/or ovary cancer. They are also useful in the gene therapy of breast and ovarian cancer.

Sequence 491 AA;

Query Match 31.7%; Score 1554; DB 22; Length 491;  
Best Local Similarity 99.3%; Pred. No. 6.6e-89;  
Matches 272; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 584 MIFQENGVYQYVVISPPPILENPTPEPPVQLOPEILLVEPLAPAPARTPTGLQ 643  
Db 1 mifqenpgvyqyvvispppiilenptpeppvqlpeillveplapapartptgtlq 60  
QY 644 ROVRIPQAPPHPTPLGSPAAYWKRVGHSACSGKGVWRPFLICISRESGEELDER 703  
Db 61 rqvripqappphprtplgspaaywkrvghsaacsqkgwrflicisresgeelider 120  
QY 704 SCAAGARPAPPECHGTPCPYPWEAGTWTSCRSCTGTOHRLQCRQERGGGSSVPP 763  
Db 121 scaagarpaspechgtcpypweagwtscrsctgptdhrqlcrqerfgggssvpp 180  
QY 764 ERCGHLPRNITQSCRLCQHWEGVSWQSVRCGRQSRQVRCVGNNGDEVSEQC 823  
Db 181 ercghlprnitqscqlrcqhwegvswqsvrcgrqsrqvcvgnngdevseqc 240  
QY 824 ASGPQPSREACDMGPTTAWFHSDDWSKVSPE 857  
Db 241 asgpppsreacdmgpttawfhsdwsksae 274

RESULT 2

AAE09713  
ID AAE09713 standard; Protein; 260 AA.

AC AAE09713;

XX 22-NOV-2001 (first entry)

DE Human gene 9 encoding novel protein HUCMO06, SEQ ID NO:60.

Human; cytostatic; gene therapy; inflammatory disorder; neural disorder; Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia; reproductive disorder; Crohn's disease; pulmonary disorder; cancer; myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour; haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis; anti-sense therapy; endocrine disorder; leukaemia.

OS Homo sapiens.

XX Key Location/Qualifiers

PH FT Misc-difference 235

FT /label= Unknown

FT /note= "Encoded by TGN"

XX WO200155202-A1.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01325.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.

PR	02-OCT-2000;	2000US-0237039.	PS	Claim 11; SEQ ID NO 60; 469pp; English.
PR	13-OCT-2000;	2000US-0237040.	XX	AAAD16750-AAAD16775 represent cDNAs corresponding to novel human protein
PR	13-OCT-2000;	2000US-0239935.	CC	genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777
PR	13-OCT-2000;	2000US-0239937.	CC	-AAAD16780 represent novel human genomic DNA fragments. The novel proteins
PR	20-OCT-2000;	2000US-0240960.	CC	and their DNAs are useful for diagnosing, treating, preventing and/or
PR	20-OCT-2000;	2000US-0241221.	CC	prognosing inflammatory disorders (bursitis or tendonitis); neural
PR	20-OCT-2000;	2000US-0241785.	CC	disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system
PR	20-OCT-2000;	2000US-0241786.	CC	disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);
PR	20-OCT-2000;	2000US-0241787.	CC	muscular disorders; reproductive disorders; gastrointestinal disorders
PR	20-OCT-2000;	2000US-0241808.	CC	(malabsorption syndrome, Crohn's disease); pulmonary disorders;
PR	20-OCT-2000;	2000US-0241809.	CC	cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);
PR	20-OCT-2000;	2000US-0241826.	CC	renal disorders (glomerulonephritis, nephrotic syndrome); cancerous
PR	01-NOV-2000;	2000US-0244617.	CC	disease and conditions (breast cancer); hyperproliferative disorders
PR	08-NOV-2000;	2000US-0244774.	CC	(leukaemia, hyperplasia); tumours; foetal and developmental
PR	08-NOV-2000;	2000US-0244775.	CC	abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,
PR	08-NOV-2000;	2000US-0246476.	CC	asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine
PR	08-NOV-2000;	2000US-0246477.	CC	disorders; pregnancy-related disorders and infections. The novel protein
PR	08-NOV-2000;	2000US-0246478.	CC	DNA is useful in gene therapy and anti-sense therapy. The proteins can
PR	08-NOV-2000;	2000US-0246523.	CC	also be used to aid wound healing and epithelial cell proliferation,
PR	08-NOV-2000;	2000US-0248524.	CC	to prevent skin aging due to sunburn, to maintain organs before
PR	08-NOV-2000;	2000US-0248525.	CC	transplantation, for supporting cell culture of primary tissues, to
PR	08-NOV-2000;	2000US-0248526.	CC	regenerate tissues, to identify their cognate ligands or binding
PR	08-NOV-2000;	2000US-0248527.	CC	
PR	08-NOV-2000;	2000US-0246528.	CC	
PR	08-NOV-2000;	2000US-0246532.	CC	
PR	08-NOV-2000;	2000US-0246609.	CC	
PR	08-NOV-2000;	2000US-0246610.	CC	
PR	08-NOV-2000;	2000US-0246611.	CC	
PR	08-NOV-2000;	2000US-0246613.	CC	
PR	17-NOV-2000;	2000US-0249207.	QY	708 GARPPASPEPCHTGTPCPYWEAGWTSCRSQGPCTQHRLQCRQFGGGSSVPPERC 767
PR	17-NOV-2000;	2000US-0249208.	DB	6 GCQAPSPCHTGTCPYWEAGWTSCRSQGPCTQHRLQCRQFGGGSSVPPERC 65
PR	17-NOV-2000;	2000US-0249209.	QY	768 HLPENITQSCQLRLCGHWEVGPWSQSVRCGRQSRQVRCVGNNGDEYSEQPCASGP 827
PR	17-NOV-2000;	2000US-0249210.	DB	66 HLPENITQSCQLRLCGHWEVGPWSQSVRCGRQSRQVRCVGNNGDEYSEQPCASGP 125
PR	17-NOV-2000;	2000US-0249211.	QY	828 POPPSREACDMGPTCTANFHSQWSKVSPE 857
PR	17-NOV-2000;	2000US-0249212.	DB	126 PQPPSREACDMGPTCTANFHSQWSKVSPE 155
PR	17-NOV-2000;	2000US-0249213.	RESULT	3
PR	17-NOV-2000;	2000US-0249214.	AAE09698	AAE09698 standard; Protein; 372 AA.
PR	17-NOV-2000;	2000US-0249215.	XX	AAE09698;
PR	17-NOV-2000;	2000US-0249216.	XX	22-NOV-2001 (first entry)
PR	17-NOV-2000;	2000US-0249217.	XX	Human gene 9 encoding novel protein HUCMO06, SEQ ID NO:45.
PR	17-NOV-2000;	2000US-0249218.	DE	Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;
PR	17-NOV-2000;	2000US-0249219.	XX	Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;
PR	17-NOV-2000;	2000US-0249220.	KW	autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;
PR	17-NOV-2000;	2000US-0249221.	KW	reproductive disorder; Crohn's disease; pulmonary disorder; cancer;
PR	17-NOV-2000;	2000US-0249222.	KW	myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;
PR	17-NOV-2000;	2000US-0249223.	KW	haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;
PR	17-NOV-2000;	2000US-0249224.	KW	anti-sense therapy; endocrine disorder; leukaemia.
PR	17-NOV-2000;	2000US-0249225.	XX	
PR	17-NOV-2000;	2000US-0249226.	XX	Homo sapiens.
PR	17-NOV-2000;	2000US-0249227.	OS	WO200155202-A1.
PR	17-NOV-2000;	2000US-0249228.	PN	WO200155202-A1.
PR	17-NOV-2000;	2000US-0249229.	XX	02-AUG-2001.
PR	17-NOV-2000;	2000US-0249230.	PD	17-JAN-2001; 2001WO-US01325.
PR	17-NOV-2000;	2000US-0249231.	XX	31-JAN-2000; 2000US-0179065.
PR	17-NOV-2000;	2000US-0249232.	PR	04-FEB-2000; 2000US-0180628.
PR	17-NOV-2000;	2000US-0249233.	PR	24-FEB-2000; 2000US-0184664.
PR	17-NOV-2000;	2000US-0249234.	PR	02-MAR-2000; 2000US-0186350.
PR	17-NOV-2000;	2000US-0249235.	PR	16-MAR-2000; 2000US-0189874.
PR	17-NOV-2000;	2000US-0249236.	PR	17-MAR-2000; 2000US-0190076.
PR	17-NOV-2000;	2000US-0249237.	PR	
PR	17-NOV-2000;	2000US-0249238.	PR	
PR	17-NOV-2000;	2000US-0249239.	PR	
PR	17-NOV-2000;	2000US-0249240.	PR	
PR	17-NOV-2000;	2000US-0249241.	PR	
PR	17-NOV-2000;	2000US-0249242.	PR	
PR	17-NOV-2000;	2000US-0249243.	PR	
PR	17-NOV-2000;	2000US-0249244.	PR	
PR	17-NOV-2000;	2000US-0249245.	PR	
PR	17-NOV-2000;	2000US-0249246.	PR	
PR	17-NOV-2000;	2000US-0249247.	PR	
PR	17-NOV-2000;	2000US-0249248.	PR	
PR	17-NOV-2000;	2000US-0249249.	PR	
PR	17-NOV-2000;	2000US-0249250.	PR	
PR	17-NOV-2000;	2000US-0249251.	PR	
PR	17-NOV-2000;	2000US-0249252.	PR	
PR	17-NOV-2000;	2000US-0249253.	PR	
PR	17-NOV-2000;	2000US-0249254.	PR	
PR	17-NOV-2000;	2000US-0249255.	PR	
PR	17-NOV-2000;	2000US-0249256.	PR	
PR	17-NOV-2000;	2000US-0249257.	PR	
PR	17-NOV-2000;	2000US-0249258.	PR	
PR	17-NOV-2000;	2000US-0249259.	PR	
PR	17-NOV-2000;	2000US-0249260.	PR	
PR	17-NOV-2000;	2000US-0249261.	PR	
PR	17-NOV-2000;	2000US-0249262.	PR	
PR	17-NOV-2000;	2000US-0249263.	PR	
PR	17-NOV-2000;	2000US-0249264.	PR	
PR	17-NOV-2000;	2000US-0249265.	PR	
PR	17-NOV-2000;	2000US-0249266.	PR	
PR	17-NOV-2000;	2000US-0249267.	PR	
PR	17-NOV-2000;	2000US-0249268.	PR	
PR	17-NOV-2000;	2000US-0249269.	PR	
PR	17-NOV-2000;	2000US-0249270.	PR	
PR	17-NOV-2000;	2000US-0249271.	PR	
PR	17-NOV-2000;	2000US-0249272.	PR	
PR	17-NOV-2000;	2000US-0249273.	PR	
PR	17-NOV-2000;	2000US-0249274.	PR	
PR	17-NOV-2000;	2000US-0249275.	PR	
PR	17-NOV-2000;	2000US-0249276.	PR	
PR	17-NOV-2000;	2000US-0249277.	PR	
PR	17-NOV-2000;	2000US-0249278.	PR	
PR	17-NOV-2000;	2000US-0249279.	PR	
PR	17-NOV-2000;	2000US-0249280.	PR	
PR	17-NOV-2000;	2000US-0249281.	PR	
PR	17-NOV-2000;	2000US-0249282.	PR	
PR	17-NOV-2000;	2000US-0249283.	PR	
PR	17-NOV-2000;	2000US-0249284.	PR	
PR	17-NOV-2000;	2000US-0249285.	PR	
PR	17-NOV-2000;	2000US-0249286.	PR	
PR	17-NOV-2000;	2000US-0249287.	PR	
PR	17-NOV-2000;	2000US-0249288.	PR	
PR	17-NOV-2000;	2000US-0249289.	PR	
PR	17-NOV-2000;	2000US-0249290.	PR	
PR	17-NOV-2000;	2000US-0249291.	PR	
PR	17-NOV-2000;	2000US-0249292.	PR	
PR	17-NOV-2000;	2000US-0249293.	PR	
PR	17-NOV-2000;	2000US-0249294.	PR	
PR	17-NOV-2000;	2000US-0249295.	PR	
PR	17-NOV-2000;	2000US-0249296.	PR	
PR	17-NOV-2000;	2000US-0249297.	PR	
PR	17-NOV-2000;	2000US-0249298.	PR	
PR	17-NOV-2000;	2000US-0249299.	PR	
PR	17-NOV-2000;	2000US-0249300.	PR	
PR	17-NOV-2000;	2000US-0249301.	PR	
PR	17-NOV-2000;	2000US-0249302.	PR	
PR	17-NOV-2000;	2000US-0249303.	PR	
PR	17-NOV-2000;	2000US-0249304.	PR	
PR	17-NOV-2000;	2000US-0249305.	PR	
PR	17-NOV-2000;	2000US-0249306.	PR	
PR	17-NOV-2000;	2000US-0249307.	PR	
PR	17-NOV-2000;	2000US-0249308.	PR	
PR	17-NOV-2000;	2000US-0249309.	PR	
PR	17-NOV-2000;	2000US-0249310.	PR	
PR	17-NOV-2000;	2000US-0249311.	PR	
PR	17-NOV-2000;	2000US-0249312.	PR	
PR	17-NOV-2000;	2000US-0249313.	PR	
PR	17-NOV-2000;	2000US-0249314.	PR	
PR	17-NOV-2000;	2000US-0249315.	PR	
PR	17-NOV-2000;	2000US-0249316.	PR	
PR	17-NOV-2000;	2000US-0249317.	PR	
PR	17-NOV-2000;	2000US-0249318.	PR	
PR	17-NOV-2000;	2000US-0249319.	PR	
PR	17-NOV-2000;	2000US-0249320.	PR	
PR	17-NOV-2000;	2000US-0249321.	PR	
PR	17-NOV-2000;	2000US-0249322.	PR	
PR	17-NOV-2000;	2000US-0249323.	PR	
PR	17-NOV-2000;	2000US-0249324.	PR	
PR	17-NOV-2000;	2000US-0249325.	PR	
PR	17-NOV-2000;	2000US-0249326.	PR	
PR	17-NOV-2000;	2000US-0249327.	PR	
PR	17-NOV-2000;	2000US-0249328.	PR	
PR	17-NOV-2000;	2000US-0249329.	PR	
PR	17-NOV-2000;	2000US-0249330.	PR	
PR	17-NOV-2000;	2000US-0249331.	PR	
PR	17-NOV-2000;	2000US-0249332.	PR	
PR	17-NOV-2000;	2000US-0249333.	PR	
PR	17-NOV-2000;	2000US-0249334.	PR	
PR	17-NOV-2000;	2000US-0249335.	PR	
PR	17-NOV-2000;	2000US-0249336.	PR	
PR	17-NOV-2000;	2000US-0249337.	PR	
PR	17-NOV-2000;	2000US-0249338.	PR	
PR	17-NOV-2000;	2000US-0249339.	PR	
PR	17-NOV-2000;	2000US-0249340.	PR	
PR	17-NOV-2000;	2000US-0249341.	PR	
PR	17-NOV-2000;	2000US-0249342.	PR	
PR	17-NOV-2000;	2000US-0249343.	PR	
PR	17-NOV-2000;	2000US-0249344.	PR	
PR	17-NOV-2000;	2000US-0249345.	PR	
PR	17-NOV-2000;	2000US-0249346.	PR	
PR	17-NOV-2000;	2000US-0249347.	PR	
PR	17-NOV-2000;	2000US-0249348.	PR	
PR	17-NOV-2000;	2000US-0249349.	PR	
PR	17-NOV-2000;	2000US-0249350.	PR	
PR	17-NOV-2000;	2000US-0249351.	PR	
PR	17-NOV-2000;	2000US-0249352.	PR	
PR	17-NOV-2000;	2000US-0249353.	PR	
PR	17-NOV-2000;	2000US-0249354.	PR	
PR	17-NOV-2000;	2000US-0249355.	PR	
PR	17-NOV-2000;	2000US-0249356.	PR	
PR	17-NOV-2000;	2000US-0249357.	PR	
PR	17-NOV-2000;	2000US-0249358.	PR	
PR	17-NOV-2000;	2000US-0249359.	PR	
PR	17-NOV-2000;	2000US-0249360.	PR	
PR	17-NOV-2000;	2000US-0249361.	PR	
PR	17-NOV-2000;	2000US-0249362.	PR	
PR	17-NOV-2000;	2000US-0249363.	PR	
PR	17-NOV-2000;	2000US-0249364.	PR	
PR	17-NOV-2000;	2000US-0249365.	PR	
PR	17-NOV-2000;	2000US-0249366.	PR	
PR	17-NOV-2000;	2000US-0249367.	PR	
PR	17-NOV-2000;	2000US-0249368.	PR	
PR	17-NOV-2000;	2000US-0249369.	PR	
PR	17-NOV-2000;	2000US-0249370.	PR	
PR	17-NOV-2000;	2000US-0249371.	PR	
PR	17-NOV-2000;	2000US-0249372.	PR	
PR	17-NOV-2000;	2000US-0249373.	PR	
PR	17-NOV-2000;	2000US-0249374.	PR	
PR	17-NOV-2000;	2000US-0249375.	PR	
PR	17-NOV-2000;	2000US-0249376.	PR	
PR	17-NOV-2000;	2000US-0249377.	PR	
PR	17-NOV-2000;	2000US-0249378.	PR	
PR	17-NOV-2000;	2000US-0249379.	PR	
PR	17-NOV-2000;	2000US-0249380.	PR	
PR	17-NOV-2000;	2000US-0249381.	PR	
PR	17-NOV-2000;	2000US-0249382.	PR	
PR	17-NOV-2000;	2000US-0249383.	PR	
PR	17-NOV-2000;	2000US-0249384.	PR	
PR	17-NOV-2000;	2000US-0249385.	PR	
PR	17-NOV-2000;	2000US-0249386.	PR	
PR	17-NOV-2000;	2000US-0249387.	PR	
PR	17-NOV-2000;	2000US-0249388.	PR	
PR	17-NOV-2000;	2000US-0249389.	PR	
PR	17-NOV-2000;	2000US-0249390		

PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249284.  
PR 17-NOV-2000; 2000US-0249285.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251899.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-451925/48.  
XX N-PSDB; AAD16758.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
XX medical disorders and also for testing and detection e.g. diagnosis and  
XX screening for agonists -  
XX  
XX Claim 11; SEQ ID No 45; 469pp; English.  
XX AAD16750-AAD16775 represent cDNAs corresponding to novel human protein  
XX genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777  
CC

CC -AAD16780 represent novel human genomic DNA fragments. The novel proteins  
CC and their DNAs are useful for diagnosing, treating, preventing and/or  
CC prognosing inflammatory disorders (bursitis or tendonitis); neural  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system  
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);  
CC muscular disorders; reproductive disorders; gastrointestinal disorders  
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;  
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);  
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous  
CC disease and conditions (breast cancer); hyperproliferative disorders  
CC (leukaemia, hyperplasia); tumours; foetal and developmental  
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,  
CC asthma); angiogenic disorders; diabetes; atherosclerosis; endocrine  
CC disorders; pregnancy-related disorders and infections. The novel protein  
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can  
CC also be used to aid wound healing and epithelial cell proliferation,  
CC to prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. The present sequence  
CC represents a novel human protein of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly

Query Match 17.5%; Score 858; DB 22; Length 372;  
Best Local Similarity 96.7%; Pred. No. 3.7e-45;  
Matches 145; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 708 GARPASPEPCGCHPCPPYWEAGETSCSRSCGPGTQHQRLQCRQERGGSSVPPERC 767  
D 1 : |||||  
D 6 gcqapaspepcgchgtcpbpypweagetscsrscgpgtqhrqlqcrqetggssvppercg 65  
QY 768 HLPENITQSCQLRLCGHWEVSGVSWQSVRCGRQSRQVRCVGNNGDEVSEQECASGP 827  
D 56 hlprnitqscqlrlcgwhewspwqsvrcgrgqrgrqrcvgnngdevseqcasgp 125  
QY 828 POPPREACDMGPCTATFPHSDWSKVSPE 857  
D 126 poppreacdmgpctatfphsdwskscae 155

RESULT 4  
AAB50947  
ID AAB50947 standard; peptide; 150 AA.  
XX  
AC AAB50947;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE ADAM gene #2 peptide.  
XX  
KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;  
KW ovarian.  
XX  
OS Homo sapiens.  
XX  
FN WO200073323-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 25-MAY-2000; 2000WO-US14308.  
XX  
PR 27-MAY-1999; 99US-0136388.  
PR 09-JUL-1999; 99US-0142930.  
PR 28-JAN-2000; 2000US-0176717.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;  
XX  
DR WPI; 2001-016507/02.

PT Seven nucleic acid molecules encoding ADAM polypeptides containing a  
PT disintegrin and metalloprotease domain, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
XX Disclosure; Page 17; 287pp; English.  
XX  
XX The present invention relates to seven members of the ADAMs (proteins  
XX which contain A Disintegrin And Metalloprotease domain) protein family.  
XX The ADAMs proteins and DNA may be used to treat disease, as a food  
XX additive or preservative, for chromosome identification, as probes  
XX for diagnosing a disorder related to the female reproductive system,  
XX particularly breast and/or ovarian cancer. They are also useful in the  
XX gene therapy of breast and ovarian cancer.  
XX  
XX Sequence 150 AA;

Query Match 16.7%; Score 817; DB 22; Length 150;  
Best Local Similarity 100.0%; Pred. No. 4.5e-43;  
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 435 EKQDGTLCOPGAPDICVAGRCCLSPGCDGTLGSGRRPDCGCGGDDSTCRILVSGNLTR 494  
D 2 ekvqdgclcpgapdicvagrclspgcdcgtlgsgrrrpdcgvcgddstcrilvsgnltr 61  
QY 495 GGPLGYQKILWIPAGALRLQIALRPSSNYLALRPGGRSINGNWAYDPPGSRAGGTV 554  
D 62 gplgyqkllwipagalrlqlalrpssnylalrgpggrsilngnwavdppgsraggtv 121  
QY 555 FRYNRPREEGKESLSAEGTTPQVDVY 583  
D 122 frynrpreegkessaeegtqpvdvy 150

RESULT 5  
AAB6949  
ID AAB6949 standard; Protein; 1690 AA.  
XX  
AC AAB6949;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human metalloprotease MPTS-19 protein.  
XX  
KW Metalloprotease; MPTS-19; human; thrombospondin domain; spondylitis;  
KW aggrecan associated disease rheumatoid arthritis; osteoarthritis;  
KW osteopathic; antiarthritic; antipsoriatic; antifibrotic; psoriasis;  
KW sport injury; joint trauma; fibrosis.  
XX  
OS Homo sapiens.  
XX  
PN DE10107360-A1.  
XX  
PD 06-SEP-2001.  
XX  
PF 16-FEB-2001; 2001DE-1007360.  
XX  
PR 18-FEB-2000; 2000US-184152P.  
XX  
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
PI Heller RA, Klonowski P, Zuo F;  
XX  
DR WPI; 2001-607955/70.  
DR N-PSDB; AAH49372.  
XX  
PT Novel metalloprotease containing a thrombospondin domain (MPTS protein)  
PT is useful to treat aggrecan associated disease including rheumatoid  
PT arthritis and osteoarthritis -  
XX  
XX Claim 1a; Page 33-38; 66pp; German.

CC This invention describes a novel metalloprotease containing a  
CC thrombospondin domain (MPS protein) which is useful to treat aggrean  
CC associated disease including rheumatoid arthritis and osteoarthritis.  
CC The products of the invention have osteopathic, antiarthritic,  
CC antiproliferative and antirheumatic activity. MPS may be used to identify a  
CC modulator of its activity, e.g., an agonist or antagonist. Such  
CC compounds, as well as MPS itself may be used to treat disease  
CC associated with MPS activity or aggrean degradation, particularly  
CC osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport  
CC injury, joint trauma or fibrosis. This sequence represents the human  
CC MPS-19 metalloprotease described in the method of the invention.  
XX  
SQ Sequence 1690 AA;

Query Match 15.18; Score 740.5; DB:22; Length 1690;  
Best Local Similarity 26.08; Pred. No. 3.4e-37;  
Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 39;

QY 86 PRPPRH-----PEALLPRGQGRPOTS-----PETPLRYTQSRGGRGGLRGA 129  
DB 176 parpgahphvvykraperlaqgdasapstcgvyypele----- 218  
QY 130 SHLGRETEITRAARRSKLRDIPKMGFGYGRVFPALPLHRRNRHRSPPR----- 180  
DB 219 ---rrerweqrgwrprlr-----rlhgrsvskewwetlv 253  
QY 181 SELSLISSRGEAIPSPRAEPF-----SANGSP-----QT 212  
DB 254 adaknveyhgq-----pqvesyvtlmmnavglfhdpisgnhlitvrlvledee 306  
QY 213 ELPPTELSVHTPSP-----QAEPLSPETA-----QTEVAPRTRPAP---LKH- 251  
DB 307 dlkithadnptckfwksnmkgdahplhddtailltrkldcatmnrpcetlgishv 366  
QY 252 ---HPRQAQSTEP-----SPHSLG-----EGGFRASPPQPRPSSQGNASPOV 294  
DB 367 agmcqphrscsinedtqlaftvaheghsfqhdgsndcepvgkrpfim---spql 423  
QY 295 -----AGRR-----PDP-----FPSVPRG-----RGOQG 313  
DB 424 lydaapltwscrgyitrfdrgwglclddpakdiidfpvpgvlydvshqrlqvg 483  
QY 314 -----QGPWGTGTPHPRLEPQOH-----PGAWLPL-----LSN 344  
DB 484 aysafcedmndvchtwsvgttchskidaavdgtrogenkwcslsgcevpvgrfpeavdg 543  
QY 345 GHASLSLWLFAPS-----SPIPR-----CSGESEQLRACQAPCPPEQDP 386  
DB 544 gswgswawsicrscgmvgaaeqctqptkgyrcvgerkrfrlcnlqacpagrpsf 603  
QY 387 RALQCAAFNSQEFMQLQWEPFTEVQSQRCELNCRPRGRFRFYVHTEKVDGTLQCP- 445  
DB 604 rhvqcsfdamlykgrlhtwvvp---vndvnpocelhrpaneyfaeklrдавvdgtpcyqv 661  
QY 446 -GAPDICVAGCLSPGCGILGSGRRPDCGCGGDDGTCLRLVSNLDRGPGLYGXIL 504  
DB 662 rasrdclngicnkvdcfeidsgamedrcgvchngstchtvgtfeaaq-lygyvvg 720  
QY 505 WTPAGALRLQIAQLRPSNLYALRPG-GRSLINGNNAVDPGSRAGTGTFRNRPRRE 563  
DB 721 lipagarelirdeavaeanflaersdebekyflngdqlwngdyqvagvgtftyar----- 776  
QY 564 EKGKSLSAGETTPQVDVMTFQSENPCVFYQYVISPPILENPTPEPPVPVLPQPEIL 623  
DB 777 rgnwntspgptkepvwiqlifqesngpvhyey----- 810  
QY 624 RVEPPLAPAPRPTGTLQORVIRIQPAPPHPTPLGSPAAYWKRVGHSACSASCKG 683  
DB 811 -----tihre-----agghdevp---ppvfwshygpwtckctvcgr 844  
QY 684 VWRPIFLCTSRSGEELDESRCAAGARRPASPEPCHGTCPPEYWEAGETWCSRSCGP-G 742

DB 845 vqrqnyclierqag-pvdeehcdplgrpdqqrkscsepcparwwagewqlcssscgpgg 903  
QY 743 TOHRQLQCRQEFQ-GGGSVPPERCCHLPRPNTQSCQLRL-C-CHWEVGSFWSQCSVRC 799  
DB 904 lsrravlcirsvgldeqsaleppacehlprrptetpcnrhvpccpatwvgn-wsqcsvte 962  
QY 800 GRGQRSRQVRVCVGNNDSEVSEOCASGPPPPSREACDMGPC 841  
DB 963 gegtqrnlvclntdntgvpdeag-----qpasevtctslpc 998  
RESULT 6  
AAE00934  
ID AAE00934 standard; Protein; 1686 AA.  
XX  
AC AAE00934;  
XX  
DT 04-JUL-2001 (first entry)  
DE Human 27875 ADAM-TS (a disintegrin and metalloproteinase).  
XX  
KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antinflammatory;  
KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;  
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;  
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;  
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;  
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;  
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;  
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;  
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;  
KW cardiant; tumour; thymoma; vasotropic; cytostatic; virucide.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..30  
FT /label= Signal\_peptide  
FT Modified-site 6..8  
FT /note= "Protein kinase C phosphorylation site"  
FT Protein 31..1686  
FT /label= Human\_mature\_27875\_ADAM-TS\_protein  
FT Modified-site 55..60  
FT /note= "N-myristoylation site"  
FT Modified-site 73..75  
FT /note= "Protein kinase C phosphorylation site"  
FT Domain 78..93  
FT /note= "Crystallins beta and gamma Greek key motif"  
FT Modified-site 94..97  
FT /note= "N-glycosylation site"  
FT Modified-site 110..112  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site 115..120  
FT /note= "N-myristoylation site"  
FT Modified-site 141..146  
FT /note= "N-myristoylation site"  
FT Modified-site 147..150  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site 159..162  
FT /note= "Casein kinase II phosphorylation site"  
FT Binding-site 195..197  
FT /note= "Cell attachment sequence"  
FT Modified-site 214..217  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site 214..216  
FT /note= "Protein kinase C phosphorylation site"  
FT Domain 244..259  
FT /label= Metalloproteinase\_domain  
FT Modified-site 313..315  
FT /note= "Protein kinase C phosphorylation site"  
FT Modified-site 342..345  
FT /note= "Casein kinase II phosphorylation site"  
FT Modified-site 342..344  
FT /note= "Protein kinase C phosphorylation site"

```

FT Modified-site 373..376 /note="Casein kinase II phosphorylation site"
FT Modified-site 379..384 /note="N-myristoylation site"
FT Domain 385..394 /label= Zinc_binding_domain
FT Modified-site 401..404 /note="Casein kinase II phosphorylation site"
FT Modified-site 408..411 /note="Amidation site"
FT Modified-site 479..484 /note="N-myristoylation site"
FT Domain 488..567 /label= Thrombospondin_domain
FT Modified-site 505..508 /note="Casein kinase II phosphorylation site"
FT Modified-site 513..518 /note="N-myristoylation site"
FT Domain 539..545 /note="Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site 539..544 /note="N-myristoylation site"
FT Domain 541..592 /label= Disintegrin_domain
FT Domain 542..592 /label= Thrombospondin_domain
FT Modified-site 557..562 /note="N-myristoylation site"
FT Modified-site 569..571 /note="Protein kinase C phosphorylation site"
FT Modified-site 598..600 /note="Protein kinase C phosphorylation site"
FT Modified-site 605..608 /note="Casein kinase II phosphorylation site"
FT Modified-site 614..619 /note="N-myristoylation site"
FT Modified-site 667..672 /note="N-myristoylation site"
FT Binding-site 687..692 /note="Cytochrome C family heme-binding site"
FT Modified-site 688..693 /note="N-myristoylation site"
FT Modified-site 693..696 /note="N-glycosylation site"
FT Modified-site 703..706 /note="Casein kinase II phosphorylation site"
FT Modified-site 716..721 /note="N-myristoylation site"
FT Modified-site 740..747 /note="Tyrosine kinase phosphorylation site"
FT Modified-site 765..770 /note="N-myristoylation site"
FT Modified-site 774..779 /note="N-myristoylation site"
FT Modified-site 778..781 /note="N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note="Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note="Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..998 /label= Thrombospondin_domain
FT Modified-site 950..953 /note="N-glycosylation site"
FT Modified-site 957..960

```

```

FT Modified-site 962..964 /note="Casein kinase II phosphorylation site"
FT Modified-site 971..974 /note="Protein kinase C phosphorylation site"
FT Modified-site 1005..1010 /note="N-glycosylation site"
FT Binding-site 1006..1009 /note="N-myristoylation site"
FT Modified-site 1011..1014 /note="Glycosaminoglycan attachment site"
FT Modified-site 1035..1037 /note="Casein kinase II phosphorylation site"
FT Modified-site 1039..1044 /note="Protein kinase C phosphorylation site"
FT Modified-site 1192..1195 /note="N-myristoylation site"
FT Modified-site 1263..1252 /note="Casein kinase II phosphorylation site"
FT Modified-site 1263..1268 /note="N-myristoylation site"
FT Modified-site 1308..1311 /note="N-myristoylation site"
FT Modified-site 1358..1363 /note="Casein kinase II phosphorylation site"
FT Modified-site 1370..1372 /note="N-myristoylation site"
FT Modified-site 1385..1387 /note="Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note="Protein kinase C phosphorylation site"
FT Modified-site 1412..1415 /note="Casein kinase II phosphorylation site"
FT Domain 1415..1463 /note="N-glycosylation site"
FT Modified-site 1419..1422 /label= Thrombospondin_domain
FT Modified-site 1440..1443 /note="N-glycosylation site"
FT Modified-site 1440..1442 /note="Casein kinase II phosphorylation site"
FT Domain 1466..1521 /note="Protein kinase C phosphorylation site"
FT Modified-site 1470..1473 /label= Thrombospondin_domain
FT Modified-site 1483..1486 /note="N-glycosylation site"
FT Modified-site 1483..1485 /note="Casein kinase II phosphorylation site"
FT Modified-site 1517..1522 /note="Protein kinase C phosphorylation site"
FT Modified-site 1528..1531 /note="N-myristoylation site"
FT Modified-site 1528..1530 /note="Casein kinase II phosphorylation site"
FT Modified-site 1546..1549 /note="Protein kinase C phosphorylation site"
FT Modified-site 1592..1597 /note="Casein kinase II phosphorylation site"
FT Modified-site 1599..1601 /note="N-myristoylation site"
FT Modified-site 1606..1609 /note="Protein kinase C phosphorylation site"
FT Modified-site 1606..1609 /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 1620..1622 /note="Protein kinase C phosphorylation site"
FT Modified-site 1625..1630 /note="Protein kinase C phosphorylation site"

```

Query Match 15.0%; Score 736.5; DB 22; Length 1686;  
 Best Local Similarity 26.0%; Pred. No. 6e-37;  
 Matches 245; Conservative 95; Mismatches 297; Indels 305; Gaps 40;  
 QY 86 PRPRH-----PEALLPRGQGRPQTS-----PETLPLYRQSRGGPLRGPA 129



```

Db 172 parpghaqqhvvykrqaperlaqrdsasptscgvvype-----lesr----- 215
QY 130 SHLGRETEQIEIARRSRRLRDPKFGMEGYGRVPFALPLHRNRHPRSPRPPR----- 180
Db 216 -----terweqgqrwrplr-----rlhqrsvskekwetlrv 249
QY 181 SELSLSSRGEAIPSPTRAPFP-----SANGSP-----QTELP 215
Db 250 adakmveyhgq-----pqvesyvltnmmvavglfhdpisignphltivrlvledee 302
QY 216 PTELSVHTPSP-----QAEPLSPETA-----QTEVAPTRPAP-----LRH- 251
Db 303 dlkithadntlksckwqksimkgdahplhdtailtrkdlcaamrpoetlglshv 362
QY 252 ----HPRQAQSGTEPP-----SPTHSLG-----EGGFRRASQPQRRPSSQGWASQV 294
Db 363 agmcqphrscsinedtqlpftvahlghsfghdgsndcepvgkrpfim---spql 419
QY 295 -----AGRR-----PDP-----PPSVPRG-----RQOQ 313
Db 420 lydaaplwtwscrsqyitrlfdrgwglclddpakdiidfpvpgvlydvshqcrlyg 479
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLPL-----LSN 344
Db 480 aysafcedmndvchtlwcsvgttchskldaavdgtrcgenkwcslgscvpgfrpeavdg 539
QY 345 GPHASSLSLAFAPS-----SPIPR-----CSGESEQLRACSOQAPCPFPQDP 386
Db 540 gwsawsicrsrscgmvgvsaerqctqtpkykrycvgerkrfrlcnlqacpagrpsf 599
QY 387 RALQCAAFNSQEFMGOLYQWEPFTEVQGSQRCLELNRGRFRFYVRHTEKVQDGTLCQP- 445
Db 600 rhvqcshtdamlkyqglhtwvvp--vndvnpcelhcrpaneyfaeklrдавдgtpcyqv 657
QY 446 -CAPDICVAGRLSPCGDILGSGRRPDCGVCGGDDSTCLVSNLNDRGGLYQKIL 504
Db 658 rsrdclngicnkvgocfidsamedrcgvchgngstchtvsgtfeeaag-lyyvdvg 716
QY 505 WTPAGALRLQIAQLRPSNNYLALRPG--GRSLINGNNAVDPPGSYRAGTVFRYRNRPPE 563
Db 717 lipagareiridevaeaanflalsedbekyflngwtgwiqngdvqvagttftvar----- 772
QY 564 EKGGSLSAEGTTPQVDVYIMFQENPVFYQYVVISPPPLENPTPEPPVLPQLPBIL 623
Db 773 rgnwlnltpoptkpevwiqlifqesnpgvhyey----- 806
QY 624 RVEPPLAPAPRPTPTGLQKQVRIPQMPAPHPPTPLGSPAAYKVRVGHGSACSGCKG 683
Db 807 -----tltre-----agghdevp--ppvfwyhygwpkctvtogr 840
QY 684 VWRPFLCISRESGELDERSCAGARPPASPECHGTPCPCPYWEAGWTSCRSRCPG-G 742
Db 841 vqrqnvclerqag-pvdeehcdplgrpddqgrkxscqpcparwwagewqlcssscpgg 899
QY 743 TOHROLCRQEFQ-GGGSVPPPERGCHLPRNITQSCQLRL-C-GHWEVSGSPWQCSYRC 799
Db 900 lrravlcirsvgldeqaleapacehlprptetpcnrhvcpatwavn-wsqcsvtc 958
QY 800 GRGQSRQVRQVGNNGDEVSCQACSGPQPPSRACDMGFC 841
Db 959 gegtrrnvlctndtgvpcdeaq-----qpasevtcslpic 994

RESULT 7
AAB74944
ID AAB74944 standard; Protein; 1686 AA.
XX
AC AAB74944;
XX
XX 02-JUL-2001 (first entry)
XX
DE Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.

```

```

XX
KW Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis deformans.
XX
OS Homo sapiens.
XX
PN JP2001008687-A.
XX
PD 16-JAN-2001.
XX
PF 25-JUN-1999; 99JP-0180973.
XX
PR 25-JUN-1999; 99JP-0180973.
XX
PA (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
XX WPI; 2001-285362/30.
XX N-PSDB; AAF82149.
XX
XX New metal protease and metal protease gene, for use as a drug for
XX cancers, arthritis and arthrosis deformans.
XX
XX Claim 1; Page 12-17; 31pp; Japanese.
XX
XX The present sequence represents a disintegrin and metalloprotease (ADAM)
XX type metal protease designated MDTs1, isolated from human. MDTs proteins
XX have cytosolic and antiarthritic activities. They can be used as a drug
XX for cancers, arthritis and arthrosis deformans.
XX
XX Sequence 1686 AA;

```

Query Match 15.0%; Score 735.5; DB 22; Length 1686;  
 Best Local Similarity 26.0%; Pred. No. 6.9e-37;  
 Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 40;

```

QY 86 PRPRH-----PEALLRPGQCPRTQTS-----PETLLYRTQSRGGPLRGA 129
Db 172 parpghaqqhvvykrqaperlaqrdsasptscgvvype-----lesr----- 215
QY 130 SHLGRETEQIEIARRSRRLRDPKFGMEGYGRVPFALPLHRNRHPRSPRPPR----- 180
Db 216 -----terweqgqrwrplr-----rlhqrsvskekwetlrv 249
QY 181 SELSLSSRGEAIPSPTRAPFP-----SANGSP-----QTELP 215
Db 250 adakmveyhgq-----pqvesyvltnmmvavglfhdpisignphltivrlvledee 302
QY 216 PTELSVHTPSP-----QAEPLSPETA-----QTEVAPTRPAP-----LRH- 251
Db 303 dlkithadntlksckwqksimkgdahplhdtailtrkdlcaamrpoetlglshv 362
QY 252 ----HPRQAQSGTEPP-----SPTHSLG-----EGGFRRASQPQRRPSSQGWASQV 294
Db 363 agmcqphrscsinedtqlpftvahlghsfghdgsndcepvgkrpfim---spql 419
QY 295 -----AGRR-----PDP-----PPSVPRG-----RQOQ 313
Db 420 lydaaplwtwscrsqyitrlfdrgwglclddpakdiidfpvpgvlydvshqcrlyg 479
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLPL-----LSN 344
Db 480 aysafcedmndvchtlwcsvgttchskldaavdgtrcgenkwcslgscvpgfrpeavdg 539
QY 345 GPHASSLSLAFAPS-----SPIPR-----CSGESEQLRACSOQAPCPFPQDP 386
Db 540 gwsawsicrsrscgmvgvsaerqctqtpkykrycvgerkrfrlcnlqacpagrpsf 599
QY 387 RALQCAAFNSQEFMGOLYQWEPFTEVQGSQRCLELNRGRFRFYVRHTEKVQDGTLCQP- 445
Db 600 rhvqcshtdamlkyqglhtwvvp--vndvnpcelhcrpaneyfaeklrдавдgtpcyqv 657

```

QY	446	-GAPDICVAGRCILSPCGDILGSGRRPDCGCVCVGDDSTCRLVLVSGLNLTDRGGPLGYKQL	504
Dd	447	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	658	raserdicngicknvcdfeidsgamedrcgvchngstchtvsgtfeeaeag-lgyvdvg	716
QY	505	WIPAGALRLIQAOLRESSNYLALRGPG-CRSIIINGNWAVDPGPSYRAGTGFYRYNRPPRE	563
Dd	506	: :   : : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	717	llpagareirigvaeaanflairedpekyfingwtiqwgdvyqagtftfyar----	772
QY	564	EKGESLSAEGTTPQVDVYMFIQENPCGVFYQYVISSPPPILENTPEPPVPQLQEPIEL	623
Dd	565	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	773	rgnwnltspgptkepvwiqlifgesnpghvey	806
QY	624	RVEPPLPAPRPARTGTTLQRQVRIPQMAPPHPRPRTPLGSPAAYWKRVCHSACSASCCKG	683
Dd	625	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	807	-----thre-----aghdevp--ppvfshygpwtktctvtcgrg	840
QY	684	VWRPFICISRESGEELDERSCAAGARRPASPEPHGTCCPYPWAEAGWTSCSRSCP-G	742
Dd	685	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	841	vqrqnycylcrdaq-pvdeehodplgrddqgrkcsgepcparwwagewqlcssscgpgg	899
QY	743	TQHROIQCQKEFG-GGSSVPERCCHLRPNITOSCQLRL-C-GHWEVGSFWSOCSVRC	799
Dd	744	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	900	lsrravlcirsvgldeqsaleppacehlprbtetpcnrhvcpatwavgn-wsqcvtc	958
QY	800	GROGRQVRQVCNNDEVSERCASGPQPSPREACDMGPC	841
Dd	801	: :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	959	gestqrnlctndtgypcadeq-----qpasevtcsiplc	994
RESULT 8			
AAE000913			
ID	AAE00913 standard; Protein; 1686 AA.		
XX AC	AAE00913;		
DT DT	04-JUL-2001 (first entry)		
XX XX	Human 27875 ADAM-TS protein, alternative version.		
XX KW	Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;		
KW KW	angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;		
KW KW	heart failure; cardiac hypertrophy; chronic ischaemic heart disease;		
KW KW	sickle cell disease nephropathy; urinary tract obstruction; haemostatic;		
KW KW	skeletal muscle disorder; myocardial infarction; blood vessel disorder;		
KW KW	hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;		
KW KW	gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;		
KW KW	rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;		
KW KW	rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;		
XX KW	cardiant; tumour; thymoma; vasotropic; cytostatic; virucide.		
OS OS	Homo sapiens.		
XX FH	Key Location/Qualifiers		
FT Peptide	1..30	/label= Signal_peptide	
FT Modified-site	6..8	/note= "Protein kinase C phosphorylation site"	
FT Protein	31..1686	/label= Human_mature_27875_ADAM-TS_protein	
FT Modified-site	55..60	/note= "N-myristoylation site"	
FT Modified-site	73..75	/note= "Protein kinase C phosphorylation site"	
FT Domain	78..93	/note= "Crystallins beta and gamma Greek key motif"	
FT Modified-site	94..97	/note= "N-glycosylation site"	
FT Modified-site	110..112	/note= "Protein kinase C phosphorylation site"	
FT Modified-site	115..120	/note= "N-myristoylation site"	
FT Modified-site	141..146	/note= "N-myriscotylation site"	
FT FT			

FT FT Modified-site /note= "N-myristoylation site"  
FT FT 774..779  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 778..781  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 825..868  
FT FT Domain /label= Thrombospondin\_domain  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 872..875  
FT FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"  
FT FT 901..903  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 917..920  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 949..994  
FT FT Domain /label= Thrombospondin\_domain  
FT FT 949..988  
FT FT Domain /label= Thrombospondin\_domain  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 950..953  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 957..960  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 962..964  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 971..974  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 1005..1010  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1006..1009  
FT FT Binding-site /note= "Glycosaminoglycan attachment site"  
FT FT 1011..1014  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1035..1037  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1039..1044  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1192..1195  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1263..1252  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1263..1268  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1308..1311  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1358..1363  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1370..1372  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1385..1387  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1397..1400  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1412..1415  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 1415..1463  
FT FT Domain /label= Thrombospondin\_domain  
FT FT 1419..1422  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 1440..1443  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1440..1442  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1466..1521  
FT FT Domain /label= Thrombospondin\_domain  
FT FT 1470..1473  
FT FT Modified-site /note= "N-glycosylation site"  
FT FT 1483..1486  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1483..1485  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1517..1522  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1528..1531

FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1528..1530  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1546..1549  
FT FT Modified-site /note= "Casein kinase II phosphorylation site"  
FT FT 1592..1597  
FT FT Modified-site /note= "N-myristoylation site"  
FT FT 1599..1601  
FT FT Modified-site /note= "Protein kinase C phosphorylation site"  
FT FT 1606..1609  
FT FT Modified-site /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"  
FT FT  
  
Query Watch 15.0%; Score 734.5; DB 22; Length 1686;  
Best Local Similarity 26.0%; Pred No. 8e-37;  
Matches 245; Conservative 94; Mismatches 298; Indels 305; Gaps 40;  
  
QY 86 PREPRH-----PEALLPRGGQPRPQTS-----PETPLRYTQSRGGPLRGPA 129  
Db 172 parpghaqphvvykrqaperlaqrldgssapstcgqvype-----lesr----- 215  
QY 130 SHLGREETQIRAAERSRLRDPKPGMGYGRVPPFALPLHRRNRHPRSPR----- 180  
Db 216 ----rerwegrqwrprlr-----rlhgrsvskkwvetllv 249  
QY 181 SELSLISSRGEAAIPSPTRAEPF-----SANGSP-----QTELP 215  
Db 250 adakmveyhq-----pqvesyvitimmvaglfhdpsignphtitvirlviedee 302  
QY 216 PTSLSVHTPSP-----QAEPLSPETA-----QTEVAPRTRPAP-----LRH- 251  
Db 303 dlkithhadntlksfckwksimkdgandplhdtdalllrklcaamrpeetlgshv 362  
QY 252 ----HPRAAQSGTEPP-----SPTHSLG-----EGGFFRASPPQRRSPSSGQWASQV 294  
Db 363 agmcqphrscsinedglplaftvaheighsfqldhgsgndcepvgkprflm---spql 419  
QY 295 -----AGRR-----PDP-----PPSVPRG-----RGQOG 313  
Db 420 lydaapltwrcsrqytrfldrgwglcldddppakdlidpsvppgvydvshqclqvg 479  
QY 314 -----QGPWGTGTPHGRPLEDPQH-----PGAWLP-----LSN 344  
Db 480 aysafcedmrvchtlwcvsgtlychskldaadvgtcrckenkwcslgscvpgvgrpeavdg 539  
QY 345 GPHASSLWSLPAPS-----SPIPR-----CSGESQLRACSOAPCPQPDP 386  
Db 540 gwsqswsawscscgmvgvsaerqctqtpkykgrycvgerkrifclnlqacpagxpsf 599  
QY 387 RALCAAFNSQEFMGQLYQWEPFTEVQSQRCELNCRPGRFRFYVRHTEKVDGTLQCP- 445  
Db 600 rhvgchfdamlykgqlhtwvpv--vndvnpclhcrpaneyfaeklrldavvgttcyqv 657  
QY 446 -GAPDICVAGRCLSPGCGDILGSGRRRDPDGGVCGGDDSTCLVSGNLTDRGGPLGYKIL 504  
Db 658 rasrdclcingicknvgcdfeldsgamedrcgvchngstchtvsgtfeeaeg-lgyvdivg 716  
QY 505 WIPACALRLQIAQLRPSSNYLALRPG-GRSINGNAVDPPGYSRAGGTGVRNRRPPRE 563  
Db 717 lipgareirirqevaeaanflalsedekyflongwtiqwngdyvagttfytar----- 772  
QY 564 EGKGESLSAEGTTQPDVYVMIFQENPQVYQVVISPPPILENTPPEPPVLPOLQEIL 623  
Db 773 rgnwenltspgtkepvwiglfdesnpgvhyey----- 806  
QY 624 RVEPPLAPAPARTPGTFLORQVIRPQMPAPPHPRTPLGSPAAYKRWKVSACSASCGKG 683  
Db 807 -----tjhre-----agghdevp--ppvfwshygpwtckctvtcgr 840  
QY 684 VWRPFLCISRESGEELDERSCAAGARPASPPECHGTGTPCPYWEAGETWSCSRSCGP-G 742  
Db 841 vqrqnvyclerqag-pvdeehcdplgrpddqgrkoseqpcparwwagewqlcscscqpgg 899

QY 743 TOHRQLQCRQFG-GGGSSVPPERCGLPRPNTTQSCQLRL-C-GHWEVGSWPSQCSVRC 799  
Db 900 lstravlicirsvgldeqsalpachlprptetpchrhvpapwatwvgn-wsgcswc 958  
QY 800 GRGSRQVRCVNGNDEVSEQCASGPPROPSPREACDMGPC 841  
Db 959 gegtqrnvictndtgvpcdeaq-----qpasevtcsiplc 994  
RESULT 9  
AAU01291  
ID AAU01291 standard; Protein; 634 AA.  
AC AAU01291;  
XX  
XX 18-JUL-2001 (first entry)  
XX Human Thrombospondin repeat domain protein 1, TSR1.  
XX  
XX Human; Thrombospondin repeat domain; TSRX: cancer; breast cancer;  
KW rheumatoid arthritis; ocular neovascularisation; wound healing;  
KW angiogenesis; immune associated disorder; gestational disorder;  
KW pre-eclampsia; neuronal development; immunogen; antibody; antisen-  
KW agonist; TSR1.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..634  
FT /label= Mature\_TSR1  
FT Domain 135..185  
FT /label= Complement\_precursor\_repeat\_domain  
FT Domain 140..190  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 418..473  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 477..533  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 535..591  
FT /label= Thrombospondin\_repeat\_domain  
FT Domain 596..630  
FT /label= Thrombospondin\_repeat\_domain  
XX  
XX WO200123561-A2.  
XX  
XX  
XX 05-APR-2001.  
XX  
XX 27-SEP-2000; 2000WO-US26432.  
XX  
XX 27-SEP-1999; 99US-0156217.  
XX 27-JUN-2000; 2000US-0214759.  
XX 26-SEP-2000; 2000US-0669360.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinkets RA, Vernet C, Tchernev VT, Boldog FL, Herrmann JL;  
XX  
XX WPI; 2001-266157/27.  
XX N-PSDB; AAS02432.  
XX  
XX TSRX PRO (PRO comprising thrombospondin-1 repeat) domain useful to  
PT identify molecules modulating TSRX activity or function, for treating  
PT cancer, rheumatoid arthritis and ocular neovascularisation -  
XX  
XX Claim 1; Page 6-9; 116pp; English.  
XX  
XX The sequence represents Human thrombospondin-1 repeat (TSR) domain  
CC containing protein, TSR1. Members of the TSR superfamily, TSRX proteins,  
CC include proteins responsible for cell attachment, spreading, motility,  
CC proliferation, cytoskeletal organisation, wound healing and  
CC angiogenesis. TSRX, TSRX polynucleotides and anti-TSRX antibodies are

CC useful for diagnosing, treating or preventing cancer, rheumatoid  
CC arthritis, ocular neovascularisation, wound healing, immune associated  
CC disorders and gestational diseases (e.g. pre-eclampsia). TSRX and TSRX  
CC polynucleotides can be used to identify members of the TSR superfamily,  
CC to screen for molecules which inhibit or enhance TSRX activity or  
CC function, as targets for identification of small molecules that modulate  
CC or inhibit e.g. angiogenesis or neuronal development. Also TSRX antisense  
CC molecules or other agonists are useful for detecting and treating breast  
CC cancer. TSRX proteins can be used to screen drugs or compounds that  
CC modulate TSRX activity or expression as well as to treat disorders  
CC characterised by insufficient or excessive production of TSRX or  
CC production of TSRX forms that have decreased or aberrant activity  
CC compared to TSRX wild-type. Anti-TSRX antibodies can be used to isolate  
CC TSRXs and modulate TSRX activity. Portions or fragments of TSRX cDNAs  
CC are used as polynucleotide reagents and are used for tissue typing  
CC and forensic identification.  
XX  
XX Sequence 634 AA;  
QY 176 RSPRSELSLSRGE---AIPSTPRAEPSANGSPQTELPTELSVHTPSQAEP 231  
Db 37 repgrehralcstgsepdiwlp---ra-----gcpreegrgaslgh-----l 79  
QY 232 SPETAOTEVAPTRPAPLRHRAQAQSGTEPPSTHSLGEGGFFRASPQRPSPSQGWAS 291  
Db 80 gpqevselwclsknrcitnsipaagt--lcqthti-----dkgwc 121  
QY 292 PQVAGRRPDPFPVPRG-RGQCGQ-GPWGHTGTPHPRLPEPDQHPGAWLPILLSNGPHAS 349  
Db 122 krvc---vpfgrpegvdgawgtpwgdcsrtcgvgvsssrhcdsprtigg----- 172  
QY 350 SLMSLPAPSPIPRCGSESLRACQAPCPPEQPPRALQCAAFNSQBFMGOLYQWEP 409  
Db 173 -----kylgerrrhscntddcpysqdfrevqsefodipfrgkyfwkwy 220  
QY 410 TEVQGSQRCENLRCRGRFVYVHTKVDGTLCCQCAPDICVAGCLSPGCCGILGSGR 469  
Db 221 -rgggvkacsltclaeinfyteraavvdgtpcrptdvlcvsgckhvgcdrvlgd 279  
QY 470 RPDGCGVGGDDSTCRUVSGNLTDGPGPLGQKLTIPAGALRLQLAQLRPPSSNYLALRG 529  
Db 280 redkrcvvggdsacetiegvfspagayedvvpkgswhifidlnslshlalkg 339  
QY 530 PGRSIIINGNAVDPPGYSRAGGVFRYNNRPREEGKESLSAEGTTPQVDVYMFQEE 589  
Db 340 dgesllileglpgtpphrlplagttfqlrggpdv---qsalealpinasallvmvlarte 396  
QY 590 NPGVFYQY---VSSPPPILENTPEPV-----POLQPELIRVE-PPLAPAPRPAR 637  
Db 397 lpalryrfinapiardslppyswhypwtksaqcagsgqvqavecrnqlssavaphycs 456  
QY 638 TPGTL-QRQVRIPOMAPPPHPTPLGSPAAYWKRVGHSACSASCKGKGVWRPIFLCISRES 696  
Db 457 ahsklpkracntepcpdd-----wvvgwnslcsrscdagvrsrvvcgrivs 505  
QY 697 GEE---LDERSCAAGARPPAPSPFECGHTPCPPYWEAGETWSCRSQCGPTQHRQLQCRQE 753  
Db 506 aaeekalddscapg-prppvl--eachgtpcpwaaaldwsectpscgpglrhrvvlcks- 562  
QY 754 FGGGSSVPPERCGLPRPNTTQSCQLRLC--GHWEVGSWPSQCSVRCGRQRVRCV 811  
Db 563 -adhratlppahcsaakppatmrcnrrcparwvage-wgecsaqcgvqgrsvrct 620  
QY 812 GNNGDEVSEQC 823  
Db 621 shtgq--ashec 630

Query Match 14.7%; Score 722; DB 22; Length 634;  
Best Local Similarity 30.2%; Pred. No. 1.6e-36;  
Matches 203; Conservative 76; Mismatches 291; Indels 102; Gaps 24;  
QY 176 RSPRSELSLSRGE---AIPSTPRAEPSANGSPQTELPTELSVHTPSQAEP 231  
Db 37 repgrehralcstgsepdiwlp---ra-----gcpreegrgaslgh-----l 79  
QY 232 SPETAOTEVAPTRPAPLRHRAQAQSGTEPPSTHSLGEGGFFRASPQRPSPSQGWAS 291  
Db 80 gpqevselwclsknrcitnsipaagt--lcqthti-----dkgwc 121  
QY 292 PQVAGRRPDPFPVPRG-RGQCGQ-GPWGHTGTPHPRLPEPDQHPGAWLPILLSNGPHAS 349  
Db 122 krvc---vpfgrpegvdgawgtpwgdcsrtcgvgvsssrhcdsprtigg----- 172  
QY 350 SLMSLPAPSPIPRCGSESLRACQAPCPPEQPPRALQCAAFNSQBFMGOLYQWEP 409  
Db 173 -----kylgerrrhscntddcpysqdfrevqsefodipfrgkyfwkwy 220  
QY 410 TEVQGSQRCENLRCRGRFVYVHTKVDGTLCCQCAPDICVAGCLSPGCCGILGSGR 469  
Db 221 -rgggvkacsltclaeinfyteraavvdgtpcrptdvlcvsgckhvgcdrvlgd 279  
QY 470 RPDGCGVGGDDSTCRUVSGNLTDGPGPLGQKLTIPAGALRLQLAQLRPPSSNYLALRG 529  
Db 280 redkrcvvggdsacetiegvfspagayedvvpkgswhifidlnslshlalkg 339  
QY 530 PGRSIIINGNAVDPPGYSRAGGVFRYNNRPREEGKESLSAEGTTPQVDVYMFQEE 589  
Db 340 dgesllileglpgtpphrlplagttfqlrggpdv---qsalealpinasallvmvlarte 396  
QY 590 NPGVFYQY---VSSPPPILENTPEPV-----POLQPELIRVE-PPLAPAPRPAR 637  
Db 397 lpalryrfinapiardslppyswhypwtksaqcagsgqvqavecrnqlssavaphycs 456  
QY 638 TPGTL-QRQVRIPOMAPPPHPTPLGSPAAYWKRVGHSACSASCKGKGVWRPIFLCISRES 696  
Db 457 ahsklpkracntepcpdd-----wvvgwnslcsrscdagvrsrvvcgrivs 505  
QY 697 GEE---LDERSCAAGARPPAPSPFECGHTPCPPYWEAGETWSCRSQCGPTQHRQLQCRQE 753  
Db 506 aaeekalddscapg-prppvl--eachgtpcpwaaaldwsectpscgpglrhrvvlcks- 562  
QY 754 FGGGSSVPPERCGLPRPNTTQSCQLRLC--GHWEVGSWPSQCSVRCGRQRVRCV 811  
Db 563 -adhratlppahcsaakppatmrcnrrcparwvage-wgecsaqcgvqgrsvrct 620  
QY 812 GNNGDEVSEQC 823  
Db 621 shtgq--ashec 630







Page 15

```

XX 31-MAR-2000; 2000WO-US08621.
XX PF
XX 31-MAR-1999; 99US-0127607.
XX PR
XX 02-APR-1999; 99US-0127636.
XX PR
XX 05-APR-1999; 99US-0127728.
XX PR
XX 30-MAR-2000; 2000US-0540763.
XX PR
XX (CURA-) CURAGEN CORP.
XX PA
XX Shimkets RA, Leach M;
XX PI
XX WPI: 2000-602362/57.
XX DR
XX N-PSDB; AAC75588.
XX PT
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX PS
XX Claim 11; Page 1678-1682; 5507pp; English.
XX CC
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX SQ
XX Sequence 1784 AA:
XX
XX Query Match 14.1%; Score 689; DB 21; Length 1784;
XX Best Local Similarity 30.1%; Pred. No. 5.4e-34;
XX Matches -182; Conservative 72; Mismatches 243; Indels 108; Gaps 22;
XX
XX QY 313 GQGPGWTGG--TPHGPRLEPDQHPGAWLPLLSNGPHASSLWSLFAPS----- 358
XX | : | | | | | | | | | | | | | | | | | | | |
XX Db 520 genkwcisgcvpvgfpe-----avdgwgswsawaiscscgmvgqsaerc 568
XX
XX QY 359 -SPIPR-----CGESQLRACSAQCPPPQDPDRALQCAAFNSQENFGQLYQWEPTEV 412
XX | : | | | | | | | | | | | | | | | | | | | |
XX Db 569 tqtpkykgyrcygerkrflcnlqapagrpsfrhvgcfhfdamlykqglhtwvpy--v 626
XX | | | | | | | | | | | | | | | | | | | |
XX QY 413 QGSORCELCNRGFRFYVSHTEKVQDGTLCQP--GAPDICVACRGCLSPGCDGTILGSRR 470
XX | | | | | | | | | | | | | | | | | | | |
XX Db 627 ndvnpcehcrpaneyfakkirdavdgtpcyqvaardicinglcknvgcdfeidsam 686
XX
XX QY 471 PDGGCVGGDDSTCRLVSGNLTDRGGPLGKQKLIWIPAGALRQIAQLRPSNNYLALRGP 530
XX | | | | | | | | | | | | | | | | | | | |
XX Db 687 edrcgvchgngstchtvsgtfxrrprvxygvdvglipagareirigevaeaanflarse 746
XX
XX QY 531 G-GRSIINGNWNVDPPGCSYRAGGTFFVFNRPPEEGKGESLSAEGPTQPVDVVMIFQEE 589
XX | : | | | | | | | | | | | | | | | | | | | |
XX Db 747 dpekyfngwtiqvngdyqvagttfyyar-----rgwnentlspgptkcpwi-qvpaar 801
XX
XX QY 590 NPGVYQYVSISSPPPILENTPEPPVPQLQPEILRVEPPLAPAPRPARTCTLQQRVIP 649
XX | | | | | | | | | | | | | | | | | | | |
XX Db 802 qdaagsqagvprbstlhafrbdqgvsqsy-----teqsebad-paaastsyspslkwb 855

```



Search completed: July 23, 2002, 21:04:40  
Job time: 3729 sec

OM of: US-10-041-770-2 to: N\_Geneseq\_032802.\* out\_format : pfs  
 Date: Jul 23, 2002 10:33 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=framet-p2n.model -DEV=xlh  
 -Q/cn2\_1/USPTO.spool/US10041770/runat\_22072002.153136.2727/app\_query.fasta\_1.947  
 -DB=N\_Geneseq\_032802 -CFMT=fastap -SUFFIX=ring -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
 -GAPEXT=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -HEP\_SIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US10041770 -CGN1\_1\_206 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-10-041-770-2  
 Query length: 877  
 Database: N\_Geneseq\_032802.\*  
 Database sequences: 1736436  
 Database length: 858457221  
 Search time (sec): 240.610000

score\_list:  
 Sequence Strd Orig zScore EScore Len ! Documentation  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 2371.00 1235.95 1.7e-6  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 925.00 496.89 2.5e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 925.00 493.30 3.9e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 808.50 428.09 1.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 807.50 427.87 1.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 806.50 427.06 1.9e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 804.50 426.04 2.2e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 797.50 424.99 2.5e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 796.50 424.47 2.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 796.00 423.51 3.0e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 790.00 420.27 4.6e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 789.50 421.46 4.8e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 788.00 419.94 4.8e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 780.00 416.86 7.1e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 757.50 400.80 5.6e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 753.50 401.79 4.9e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 728.00 389.30 2.5e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 714.50 384.29 4.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 714.00 376.51 1.3e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 699.50 364.62 5.8e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 698.50 374.08 1.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 676.00 361.36 8.8e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 636.00 342.51 9.9e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 626.50 339.70 1.4e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 625.00 337.26 1.9e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 618.00 328.88 5.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 596.00 320.79 1.6e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 590.50 317.12 2.6e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 589.50 314.18 3.7e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 588.00 314.68 3.5e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 588.00 314.00 3.8e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 587.00 315.58 3.1e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 581.00 313.17 4.3e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 581.00 310.57 6.0e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 572.00 307.27 9.1e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 566.00 304.11 1.4e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 556.50 298.04 3.0e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 556.50 297.99 3.0e-1  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA91178 + 551.00 301.86 1.8e-1

/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH47048 + 551.00 298.40 2.8  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAA5825 + 547.50 297.09 3.4e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAA47482 + 545.50 296.40 3.7e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ32013 + 541.00 279.36 3.3e  
 /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAAC90070 + 541.00 279.36 3.3

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAAC91178

seq\_documentation\_block:  
 ID AAC91178 standard; DNA; 2680 BP.

AC AAC91178;

DT 20-MAR-2001 (first entry)

DE ADAM gene #2.

KW ADAM; disintegrin; metalloprotease; food additive; breast cancer;

KW ovarian; SS.

OS Homo sapiens.

PN WC200073323-A2.

PD 07-DEC-2000.

PF 25-MAY-2000; 2000WO-US14308.

PR 27-MAY-1999; 99US-0136388.

PR 09-JUL-1999; 99US-0142930.

PR 28-JAN-2000; 2000US-0178717.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;

WPI; 2001-016507/02.

Seven nucleic acid molecules encoding ADAM polypeptides containing a disintegrin and metalloprotease domain, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -

XX Claim 1; Page 266; 287pp; English.

XX The present invention relates to seven members of the ADAMs (proteins which contain a Disintegrin And Metalloprotease domain) protein family. The ADAMs proteins and DNA may be used to treat disease, as a food additive or preservative, for chromosome identification, as probes for diagnosing a disorder related to the female reproductive system, particularly breast and/or ovarian cancer. They are also useful in the gene therapy of breast and ovarian cancer.

XX Sequence 2680 BP; 510 A; 864 C; 767 G; 537 T; 2 other;

alignment\_scores:  
 Quality: 2371.00 Length: 423  
 Ratio: 5.632 Gaps: 0  
 Percent Similarity: 99.527 Percent Identity: 99.527

alignment\_block:

US-10-041-770-2 x AAC91178 ..

Align seg 1/1 to: AAC91178 from: 1 to: 2680

435 GILYSLVALGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCy 451

|||||

5 GAAGAGGTCAGAGGAGGAGCCCTGTGTGAGCCCTGAGCCCTGACACTG 54

451 sValaLaGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlySerG 468

|||||

55 TGTGCTGAGCGCTGTCTGTGAGCCCGGCTGTGTGATGGATCCTTGGCTCTG 104

468 lYArgProAspGlyCysGlyValCysGlyGlyAspSerThrCys 484  
105 GCAGCGCTGATGGCTGTGAGTCTGTGGGTGATGATCTACCTGT 154  
485 ArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrG1 501  
155 CCCTTGTTCGGGAAGCTCAGTACGAGGGGCCCTGGGCTATCA 204  
501 nlySileLeuTirPileProAlaGlyAlaLeuArgLeuGlnIleAlaGlnL 518  
205 GAAGATCTTGTGGATTCAGCGGGAGCTTGGGCTCCAGATTGCCAGC 254  
518 euArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSer 534  
255 TCCGGCCTAGCTCCAACCTACCTGGACATTCGTGGCCCTGGGGCCGGTCC 304  
535 lIeIleAsnGlyAsnTirPalaValAspProProGlySerTyrArgAlaG1 551  
305 ATCATCATGGGAACCTGGCTGTGATCCCTGGTCTTACAGGGCCGG 354  
551 yGlyThrValPheArgTyrAsnArgProProArgGluGluGlyGlyG 568  
355 CGGGACCTCTTTCGATATAACCGTCCTCCAGGGAGGAGGCAAGGGG 404  
568 luserLeuSerAlaGluGlyProThrThrGlnProValAspValTyrMet 584  
405 AGAGTCTGTGGCTGAAGGCCCCACCCAGCCCTGTGGATGTCTATATG 454  
585 lIePheGlnGluGluAsnProGlyValPheTyrGlnTyrValIleSerSe 601  
455 ATCTTTCAGGAGAAACCCAGGCGCTTTTATCATGATGTCATCTCTC 504  
601 rProProProIleLeuGluAsnProThrProGluProProValProGlnL 618  
505 ACTCTCTCAATCTTGGAGACCCACCCAGAGCCCTGTGCCCCAGC 554  
618 euGlnProGluLeuArgValGluProProLeuAlaProAlaProArg 634  
555 TTCAGCGGAGATTCGAGGTGGAGCCCACTTGCCTCCGCGACCCCGC 604  
635 ProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGlnMe 651  
605 CCAGCCGGAGCCCGGACCCCTCCAGGTCAGGTGGGATCCCCAGAT 654  
651 tProAlaProProHisProArgThrProLeuGlySerProAlaAlaTyrT 668  
655 GCCCGCCCGCCCATCCAGACACCCCTGGGGTCTCCAGTGGGTACT 704  
668 rPlyArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyVal 684  
705 GGAACGAGTGGGACACTGTGCATGCTCAGCGTCTCGGGGAAGGTGTC 754  
685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluLeuAs 701  
755 TGCGCCCATTTTCTCTGTCATCTCCGTGAGTCGGGAGAGAACTGGA 804  
701 pGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluProC 718  
805 TGAACGACGCTGCGGGGGTCCAGGCCCCAGCCCTCCCTGAAACCT 854  
718 yHisGlyThrProCysProProTyrTirPgluAlaGlyGluTirPthrSer 734  
855 GCCAGGCAACCCATGCCCCCATCTACTGGAGCTGGCAGTGGACATCC 904  
735 CysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysAr 751  
905 TGCAGCGCTCTGTGGCCCGGACCCAGCACCCGCTGCTGCTGCTGCG 954  
751 gGlnGluPheGlyGlyGlySerSerValProProGluArgCysGlyH 768  
955 GCAGGAATTTGGGGGGGTGGCTCTCTCGTGGCCCCCGGAGCGCTGTGAC 1004

768 isLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGly 784  
1005 ATCTCCCGCCGCCAACATCACCCAGTCTTGCCAGCTGCCCTCTGTGC 1054  
785 HisTrpGluValGlySerProTrpSerGlnCysSerValArgCysGlyAr 801  
1055 CATTGGGAAGTTGGCTCTCTTGGAGCCAGTGTCCGTGGGTGGCGCG 1104  
801 gGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluV 818  
1105 GGGCAGAGAGCGCGCAGGTTGGCTGTGTGGGAACAACGGTGTATGAAG 1154  
818 alSerGluGlnGluCysAlaSerGlyProGlnProGlnProProSerArgGlu 834  
1155 TGACGAGCAGAGAGTGTGCTGAGCCGCCACAGCCGCCACACAGAG 1204  
835 AlaCysAspMetGlyProCysThrThrAlaTirPheHisSerAspTrpSe 851  
1205 GCCTGTGACATGGGCGCTGTACTACTGCTGCTGCTGCTGCTGCTGCTG 1254  
851 rSerLysValSerProGlu 857  
1255 CTCCAAGTGTCTAGCCGAG 1273  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAD16773  
seq\_documentation\_block:  
ID AAD16773 standard; cDNA; 1020 BP.  
XX  
AC AAD16773;  
XX  
DT 22-NOV-2001 (first entry)  
XX  
DE Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:34.  
KW Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;  
Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;  
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;  
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;  
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;  
KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;  
KW anti-sense therapy; endocrine disorder; leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 65..847  
FT /\*tag= a  
FT /product= "Human novel protein"  
FT /transl\_except= (pos:767..769, aa:Xaa)  
FT /note= "CDS does not include start codon; Xaa is an  
FT unknown amino acid"  
FT /partial  
XX  
WO200155202-A1.  
XX  
PN 02-AUG-2001.  
XX  
PD 17-JAN-2001; 2001WO-US01325.  
XX  
PF 31-JAN-2000; 2000US-0179065.  
XX  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.

PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	14-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246611.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226681.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226688.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000;	2000US-0228099.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229359.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-02321968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	26-SEP-2000;	2000US-0234998.	XX	(HUMA-) HUMAN GENOME SCI INC.	
PR	26-SEP-2000;	2000US-0235484.	PA	Rosen CA, Barash SC, Ruben SM;	
PR	27-SEP-2000;	2000US-0235834.	XX	WPI: 2001-451925/48.	
PR	27-SEP-2000;	2000US-0235836.	XX	P-PSDB: AAE09713.	
PR	29-SEP-2000;	2000US-0236327.	DR	Isolated polypeptide for treating, preventing and/ or prognosing	
PR	29-SEP-2000;	2000US-0236367.	DR	medical disorders and also for testing and detection e.g. diagnosis and	
PR	29-SEP-2000;	2000US-0236368.	XX	screening for agonists -	
PR	29-SEP-2000;	2000US-0236369.	XX	Claim 1; SEQ ID No 34; 469pp; English.	
PR	29-SEP-2000;	2000US-0236370.	XX	ADL6750-ADL6775 represent cDNAs corresponding to novel human protein	
PR	02-OCT-2000;	2000US-0236802.	CC	genes, and AAE09690-AAE09715 represent the proteins they encode. AAE09715	
PR	02-OCT-2000;	2000US-0237037.	CC	-AAE09715 represent novel human genomic DNA fragments. The novel proteins	
PR	02-OCT-2000;	2000US-0237038.	CC	and their DNAs are useful for diagnosing, treating, preventing and/or	
PR	02-OCT-2000;	2000US-0237039.	CC	prognosing inflammatory disorders (bursitis or tendonitis); neural	
PR	02-OCT-2000;	2000US-0237040.	CC	disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system	
PR	13-OCT-2000;	2000US-0239935.	CC	disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);	
PR	20-OCT-2000;	2000US-0240960.	CC	muscular disorders; reproductive disorders; gastrointestinal disorders	
PR	20-OCT-2000;	2000US-0241221.	CC	(malabsorption syndrome, Crohn's disease); pulmonary disorders;	
PR	20-OCT-2000;	2000US-0241785.			
PR	20-OCT-2000;	2000US-0241786.			
PR	20-OCT-2000;	2000US-0241787.			
PR	20-OCT-2000;	2000US-0241808.			
PR	20-OCT-2000;	2000US-0241809.			

```
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous
CC disease and conditions (breast cancer); hyperproliferative disorders
CC (leukaemia, hyperplasia); tumours; foetal and developmental
CC abnormalities; haematopoietic disorders; respiratory disorders (rhinitis,
CC asthma); angioneurotic disorders; diabetes; atherosclerosis; endocrine
CC disorders; pregnancy-related disorders and infections. The novel protein
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can

alignment_scores:
Quality: 925.00      Length: 172
Ratio: 5.572        Gaps: 0
Percent Similarity: 96.512  Percent Identity: 96.512

alignment_block:
US-10-041-770-2 x AAD16773 ..
Align seg 1/1 to: AAD16773 from: 1 to: 1020

686  ArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAspQ1 702
      |||||
15  CGTCCGCCACGCGTCCGATCCCGTGAGTCGGGAGAGAACTGGATGA 64
      |||||
702  uArgSerCysAlaAlaGlyAlaArgProAlaSerProGluProCysH 719
      |||||
65  ACGCAGCTGTGCCGCGGTGCCAGGCC.CCAGCCTCCCTGAACCTGCC 113
      |||||
719  isGlyThrProCysProProTyrTirPgluAlaGlyGluTirPThrSerCys 735
      |||||
114  ACGGCACCCCATGCCCCCATCTACTGGAGGCTGGCGAGTGGACATCTGC 163
      |||||
736  SerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgG1 752
      |||||
164  AGCCGCTCTGTGGCCCGCGCACCCAGCACCGCCAGCTGCAGTGGCCGCA 213
      |||||
752  nGluPheGlyGlyGlySerSerValProGluArgCysGlyHisL 769
      |||||
214  GGAATTTGGGGGGGTGGCTCTCTCGTGGCCCGAGCGCTGTGGACATC 263
      |||||
769  euProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHis 785
      |||||
264  TCCCGCGGCCAACATCACCAGCTCTTGCAGCTTGCAGCTGGCGCTGTGCCCAT 313
      |||||
786  TrpGluValGlySerProTyrSerGlnCysSerValArgCysGlyArgG1 802
      |||||
314  TGGGAATTTGGCTCTCTCTGGAGCCAGTGTCTCGTGGGTGGCGCGGG 363
      |||||
802  yGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValS 819
      |||||
364  CCAGAGAAGCGCGCAGGTTCGCTGTGTGGGAACAATGGTGAAGTGA 413
      |||||
819  erGluGlnGluCysAlaSerGlyProGlnProGlnProSerArgGluAla 835
      |||||
414  GCGAGCAGGAGTGGCTCAGGCCCCCAGCCCCCAGCCCCCAGAGAGGCC 463
      |||||
836  CysAspMetGlyProCysThrAlaTirPheHisSerAspTrpSerSe 852
      |||||
464  TGTGACATGGGCGCTCTACTACTGTCTGTCTTCCACAGCAGCTGGAGCTC 513
      |||||
852  rLysValSerProGlu 857
      |||||
514  CAAGTGTCTCAGCCGAG 529

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: AAD16758
seq_documentation_block:
ID AAD16758 standard; cDNA; 1998 BP.
XX
AC
XX
XX
DT 22-NOV-2001 (first entry)
XX
DE Human novel protein-encoding gene 9 cDNA clone HUCMO06, SEQ ID NO:19.
```

```
XX Human; cytostatic; gene therapy; inflammatory disorder; neural disorder;
KW Parkinson's disease; Alzheimer's disease; immune system disorder; AIDS;
KW autoimmune disease; rheumatoid arthritis; muscular disorder; ischaemia;
KW reproductive disorder; Crohn's disease; pulmonary disorder; cancer;
KW myocardial infarction; glomerulonephritis; nephrotic syndrome; tumour;
KW haematopoietic disorder; rhinitis; asthma; diabetes; atherosclerosis;
KW anti-sense therapy; endocrine disorder; leukaemia; ss.
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 56..1174
FT /*tag= a
FT /product= "Human novel protein"
FT /notes= "CDS does not include start codon"
FT /partial
XX
XX WO200155202-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01325.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184564.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
```

```
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 23-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249255.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451925/48.
XX P-PSDB; AAE09698.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT medical disorders and also for testing and detection e.g. diagnosis and
PT screening for agonists -
XX
XX Claim 1; SEQ ID No 19; 469pp; English.
XX
CC AAD16750-AAD16775 represent cDNAs corresponding to novel human protein
CC genes, and AAE09690-AAE09715 represent the proteins they encode. AAD16777
CC -AAD16780 represent novel human genomic DNA fragments. The novel proteins
CC and their DNAs are useful for diagnosing, treating, preventing and/or
CC prognosing inflammatory disorders (bursitis or tendonitis); neural
CC disorders (e.g. Parkinson's disease, Alzheimer's disease); immune system
CC disorders; AIDS; autoimmune diseases (e.g., rheumatoid arthritis);
CC muscular disorders; reproductive disorders; gastrointestinal disorders
CC (malabsorption syndrome, Crohn's disease); pulmonary disorders;
CC cardiovascular disorders (myocardial infarction, ischaemia, arrhythmias);
CC renal disorders (glomerulonephritis, nephrotic syndrome); cancerous
CC (leukaemia, hyperplasia); tumours; foetal and developmental
CC abnormalities; haematopoietic disorders; diabetes; atherosclerosis; endocrine
CC asthma; angiogenic disorders; diabetes; and infectious. The novel protein
CC disorders; pregnancy-related disorders and anti-sense therapy. The proteins can
CC DNA is useful in gene therapy and anti-sense therapy. The proteins can
CC also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before

alignment_scores:
Quality: 925.00 Length: 172
Ratio: 5.572 Gaps: 0
Percent Similarity: 96.512 Percent Identity: 96.512

alignment_block:
US-10-041-770-2 x AAD16758 ..
Align seg 1/1 to: AAD16758 from: 1 to: 1998

686 ArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAspG1 702
|||||
6 CQTCCGCCACACGCTCCGATCTCCGTGAGTCGGAGAGAACTGGATGA 55
|||||

702 uArgSerCysAlaAlaGlyAlaArgProAlaSerProGluProCysH 719
|||||
56 AGCAGAGTGTGCGCGGTGCCAGGCC.CCAGCCTCCCTGAACCCCTGCC 104
|||||

719 IsGlyThrProCysProProTyrTrpGluAlaGlyGluTrpThrSerCys 735
|||||
105 ACGGACCCCATGCCCCCATCTGGAGGTGGCGGAGTGACATCTTGC 154
|||||

736 SerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgG1 752
|||||
155 AGCCGCTCTGTGGCCCGGACCCAGCCAGCCAGCTGCAGTGCCTGCCGA 204
|||||
```

```

752 nGluPheGlyGlyGlySerValProGluArgCysGlyHisL 769
205 GGAATTTGGGGGGGTGGCTCTCGTGGCCCGGAGCGCTGTGACATC 254
769 euProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHis 785
255 TCCCGCGGCCAACATCACCCAGTCTTGCCAGCTGGCGCTCTGTGGCCAT 304
786 TrpGluValGlySerProTrpSerGlnCysSerValArgCysGlyArgG1 802
305 TGGGAAGTTGGCTCTCCTTGGAGCAGTCTCCGTGGCGTGGCGCGGG 354
802 yGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValS 819
355 CCAGAGAACCGCGCAGGTGCTGCTGTGGGAACAATGTTGATGAAGTGA 404
819 erGluGlnGluCysAlaSerGlyProGlnProGlnProSerArgGluAla 835
405 GCGAGCAGGAGTGTGGTCAGGCCCGCCACAGCCCCCAGCAGAGAGGCC 454
836 CysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTrpSerSe 852
455 TGTGACATGGGCGCTGTACTACTGCTGTGTCCACAGCGACTGGAGCTC 504
852 rlyValSerProGlu 857
505 CAAGTGTCTCAGCGCGAG 520

```

seq\_name: /SID81/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: AAD04494

seq\_documentation\_block:

ID\_ AAD04494 standard; cdna; 5353 BP.

XX AC AAD04494;

XX 04-JUL-2001 (first entry)

XX Human 27875 ADAM-TS (a disintegrin and metalloproteinase) cDNA.

XX Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;  
 KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;  
 KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;  
 KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;  
 KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;  
 KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;  
 KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;  
 KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;  
 KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;  
 KW cardiant; tumour; thymoma; vasotropic; cytostatic; virucide; ss.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 36..5036

XX /tag= a

XX /product= "Human 27875 ADAM-TS protein"

XX sig\_peptide 36..125

XX /tag= b

XX mat\_peptide 126..5093

XX /tag= c

XX /product= "Human mature 27875 ADAM-TS protein"

XX WO200131034-A1.

XX PD 03-MAY-2001.

XX PF 25-OCT-2000; 2000WO-US293380.

XX PR 25-OCT-1999; 99US-0426282.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Kapeller-Libermann R, White D;

XX WPI: 2001-300513/31.

DR P-PSDS; AAR00934.

XX

PT Novel isolated polypeptide, 27875, a human ADAM-TS (a disintegrin and  
 PT metalloproteinase) useful for diagnosis and treatment of disorders of  
 PT bone, lung, heart, skeletal muscle such as osteoporosis, emphysema,  
 PT angina

XX

PS Claim 7; Fig 1; 133pp; English.

XX

CC The present sequence is a cDNA encoding 27875 protein, a human ADAM-TS  
 CC (a disintegrin and metalloproteinase). Metalloproteinase is a  
 CC proteolytic enzyme involved in protein maturation, protein degradation,  
 CC tumour growth, metastasis and angiogenesis. Nucleotides encoding 27875,  
 CC 27875 protein and its antibodies are useful for preventing, diagnosing  
 CC and treating 27875 metalloproteinase-related disorders. These disorders  
 CC include, disorders involving the lung such as congenital anomalies  
 CC bronchial asthma, Goodpasture's syndrome, pulmonary alveolar proteinosis,  
 CC disorders involving heart such as heart failure, cardiac hypertrophy,  
 CC angina pectoris, myocardial infarction, chronic ischaemic heart disease,  
 CC disorders involving the skeletal muscle include tumours such as  
 CC rhabdomyosarcoma, disorders involving blood vessel such as hypertension,  
 CC atherosclerosis, vasculitis associated with other disorders, disorders  
 CC involving the testis and epididymis such as nonspecific epididymitis and  
 CC orchitis, gonorrhea, mumps, tuberculosis and syphilis, spermatocytic  
 CC seminoma, disorders involving the kidney such as membranoproliferative  
 CC glomerulonephritis, necrotising glomerulonephritis, renal artery  
 CC stenosis, chronic glomerulonephritis, sickle cell disease nephropathy,  
 CC urinary tract obstruction, disorders of the bone such as achondroplasia,  
 CC osteoporosis, Paget's disease, rickets, osteomalacia, osteonecrosis,  
 CC tuberculous osteomyelitis, osteoma, osteosarcoma, metastatic tumours,  
 CC and disorders involving thymus such as thymic cysts, thymic hypoplasia,  
 CC thymomas, and Hodgkin's disease. 27875 sequence is used in gene therapy.  
 CC Note: This sequence SEQ.ID.NO.2 is stated as being similar to the  
 CC sequence shown in Page 115-123 (See AAD04475). However these sequences  
 CC differ at several locations.

XX

SQ Sequence 5353 BP; 960 A; 1833 C; 1649 G; 911 T; 0 other;

alignment\_scores:

Quality: 808.50 Length: 959

Ratio: 1.817 Gaps: 41

Percent Similarity: 46.403 Percent Identity: 28.050

alignment\_block:

US-10-041-770-2 x AAD04494 ..

Align seg 1/1 to: AAD04494 from: 1 to: 5353

15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24

631 GTGCTCCAAGCACCTGTGGAGTGGCAAGTGTACCCAGAGCTGGAGTCTCGA 680

24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35

681 CGGGAGCGTGTGGAGCAGCGGCAGTGGCGGCGCCACGGCTGAGCGG 730

35 hrProThrGluGluGlnGlyProGluGlyValTrpGlyProTrp... 50

731 TCTACACCAGCGGTGGTGCACAAAGAGAAGTGGGTGGAGACCTGGTAG 780

50 ..... 50

781 TAGCTGATGCCAAATGTGGAGTACCAGGACAGCGCGAGTTGAGAGC 830

51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61

831 TATGTGCTGACCATCATGAACATGTGGCTGTGGCTGTTCATGACCCAG 880

61 lyValGlyVal.....Gln 65

```

881  CATTGGGAACCCATCCACATCCACATTGTGGCTGTCTGCTGGAAG 930
      :::::||||:
66  ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro 80
      ||| ||||| ::| ||||| |||
931  ATGAGGAGGAGCACTAAGATTCAGCCACCATGCAAGAACACCCCTGAAG 980
      ||| ||||| ::| ||||| |||
81  .....SerLeuProLeuP 85
      :::::||||:
981  AGCTTCTCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCATCC 1030
      ||| |||||
85  roProArgProProArgHisProGluAlaLeuProArgGlyGlnGly 101
      ||
1031  CCTGCACCATG..... 1041
102  ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118
      ||||| |||||
1042  ACACCTGCCATCTGCTACCAGAA.....AGGACCTGTGT 1076
118  rArgGlyArgGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
      ::| ||||| |||||
1077  GCAGCCATGAACCGCCCTGTGAGACCCCTGGGACTGTCCCATGTGGCGG 1126
135  luGluThrGlnGlnIleArgAlaAlaArgSerArgLeuArgAspPro 151
      :::::||||:
1127  CATGTGCCAGCGCCAGCCAGCTGCAGCATCAACAGGACA..... 1167
152  IleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuProLe 168
1167 ..... 1167
168  uHisArgAsnArgArgHisProArgSerProArgSerGluLeuSerL 185
      ||| |||||
1168  ...CGGGCTCGCGTGCCTTCACTAGCCACGAGCTCGGGCACAG 1213
185  euIleSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
      ::||| |||||
1214  TTTTGCATTCAGCATGACGAGCGGCAATGACTGTGAGCCCGTTGGGA 1263
202  GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
      :::::||||:
1264  AGCAGCTTTCATCATGTCTCCACAGCTCCTGTACGACCCGCTCCCTC 1313
216  .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
      ||| |||||
1314  ACCTGTCCCGTGCAGCGCCAGTATATACCAAGTTCTTTGACCGTGG 1363
232  exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
      ::||| |||||
1364  GTGGGCGCTGTGCTGGACGACCCCTCTGCCAAGGACATATCGACTTCC 1413
249  LeuArgHisHis.....ProArgAlaGlnAlaSe 258
      ||| |||||
1414  CCTCGTGCCACCTGGCGTCTCTATGATGTAAGCCACCATGTCGCGCTC 1463
258  rGlyThrGluProProSerPro.....T 266
      ||||| |||
1464  CAGTACGGGGCGCTACTCTCCCTCTCCGAGGACATGATTAATGTCTGCCA 1513
266  hrHisSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArg 282
      ||||| |||||
1514  CACACTCTGGTCTGTGGGA.....CCACTGTCT 1545
283  ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297
      ||||| |||
1546  ACTCCAAGCTGTATGACGCGGTGAGCGGACCGCGGTGTG...GGGAGAT 1592
298 .....ArgProAspProPheProSerValProArgG 308
1593  AAGTGTGTCTAGTGGGAGTGCCTACCGTGGGCTTCGGGCCCGAGS. 1641
308  lYargGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321
      ||||| |||||
```



```
610 hrProGluProProValProGlnLeuGlnProGluIleLeuArgValGlu 626
2453 .....
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG1 643
2454 .....ACCATCCA 2461
643 nArgInValArgIleProGlnMetProAlaProProHisProArgThrP 660
2462 CAGGGAG.....GCAGGTGGCCACGACGAGTCC 2490
660 rLeuGlySerProAlaAlaTrpTrpLysArgValGlyHisSerAlaCys 676
2491 CG.....CCGCCGTGTCTCCTGGCATTATGGCCCTGGACCAAGTGC 2534
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693
2535 ACAGTCACCTCGGCAGAGGTGTGCAGAGGAGAAATGTGTACTGCTTGG 2584
693 rArgGluSerGlyGluGluLeuAspLysArgSerCysAlaAlaGlyAla 710
2585 CGCGCAGCAGGG...CCCGTGGAGGAGGACACTGTGACCCCTGGGCC 2631
710 rgProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726
2632 GGCCTGATGACCAACAGAGGAAGTGCACGCGACGACCCCTGCCCTGCC 2681
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...G1 742
2682 TGGTGGCAGGTGAGTGGCAGCTGTGCTCCAGCTCCTCGGGCCCTGGGG 2731
742 yThrGlnHisArgGlnLeuGlnCysArgGlnGlnPheGly...GlyGly 758
2732 CQCTCCCGCGCGCGCTGCTGTGATCCGAGCGTGGGCTGGATGAGC 2781
758 LysSerValProProGluArgCysGlyHisLeuProArgProAsnIle 774
2782 AGAGCGCCTGGAGCCACCGCCCTGTGAACACCTCCCGCCCTACT 2831
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValG1 789
2832 GAAACCCCTTGCAACCGCATGTACCTGTCCGGCCACCTGGCTGTGG 2881
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSera 806
2882 GAAC...TGGTCTCAGTCTCAGTGACATGTGGGAGGCGCACTCAGCC 2928
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822
2929 GAAATGCTCTGCAACCAATGACACCGGTGTCCCTGTGACGAGGCCAG 2978
823 CysAlaSerGlyProProGluProProSerArgGluAlaCysAspMetG1 839
2979 .....CAGCCAGCCAGCAAGTCACTGCTCTCTGCC 3010
839 yProCysThr.ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854
3011 ACTGTGTCGGTGGCCCTGGGCACAC.....TGGCCCTCAAGGCTCAG 3054
855 SerProGluProProAlaIleSer 862
3055 GCAGCGGTCTCTCCAGCCAGCAGC 3078
```

seq\_name: /SID1/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF82149

seq\_documentation\_block:

ID: AAF82149 standard; cDNA; 5061 BP.

XX

AC

XX

DT 02-JUL-2001 (first entry)

```
XX Human ADAM type metal protease MDTs1 encoding cDNA SEQ ID NO:2.
DE
XX
KW Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis deformans; ss.
OS Homo sapiens.
PH Key Location/Qualifiers
FT CDS 1..5061
FT /tag= a
FT /product= "MDTs1"
FT /note= "a disintegrin and metalloprotease (ADAM)
FT type metal protease"
XX
PN JP2001008687-A.
XX
XX 16-JAN-2001.
XX
XX 25-JUN-1999; 99JP-0180973.
XX
XX 25-JUN-1999; 99JP-0180973.
XX
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
XX WPI: 2001-285362/30.
XX P-PSDB; AAF82149.
XX
XX New metal protease and metal protease gene, for use as a drug for
XX cancers, arthritis and arthrosis deformans
XX
XX Example 1; Page 17-19; 31pp: Japanese.
XX
XX The present sequence encodes a disintegrin and metalloprotease (ADAM)
XX type metal protease designated MDTs1, isolated from human. MDTs proteins
XX have cytosolic and antiarthritic activities. They can be used as a drug
XX for cancers, arthritis and arthrosis deformans.
XX
XX Sequence 5061 BP; 878 A; 1746 C; 1587 G; 850 T; 0 other;
```

```
alignment_scores:
  Quality: 807.50      Length: 959
  Ratio: 1.815        Gaps: 41
  Percent Similarity: 46.403  Percent Identity: 28.050

alignment_block:
US-10-041-770-2 x AAF82149 ..
Align seg 1/1 to: AAF82149 from: 1 to: 5061
```

```
15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
596 GTGCTCCACAGCAGCTGGAGTGCAAGTGTAACCCAGAGCTGGAGTCTCGA 645
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
646 CGGAGCGGTTGGGAGCAGCGGACGAGTGGCGCGCCAGCGCTGAGCG 695
35 hrProThrGluGluGlyGlnGlyProGluGlyValTrpGlyProTrp... 50
696 TCTACACCGGTGCTGTCAGCAAGAGAAAGAGTGGGTGGAGACCTGGTAG 745
50 ..... 50
746 TAGCTGATGCCAAATGGTGGAGTACCACGACGACGCGGAGTTGAGAGC 795
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
796 TATGTGCTGACCATCATGAATGGTGGCTGGCCTTTTCATGACCCAG 845
61 lyValGlyVal.....Gln 65
```

```

846 CATTTGGAAACCCATCCACATCACCATTGTGGCGCTGGTCTGGTGAAG 895
      ::::|||||
66  ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro. .... 80
      ||| ||||| :::: ||||| |||
896 ATGAGGAGGAGGACCTAAGATCAGCCACCATGACAGACACACCTGAAG 945
81  .....SerLeuProLeuP 85
      ::::|||||
946 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCCATCC 995
85  roProArgProProArgHisProGluAlaLeuProArgGlyGlnGly 101
      |||
996 CTGCGACCATG ..... 1006
102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118
      :::: ||||| |||||
1007 ACACCTGCCATCTGCTCCACAGAA ..... AGGACCTGTGT 1041
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
      :::: ||||| |||||
1042 GCAGCATGACCGCCCTGTGTGAGCCCTGGGACTGTCCCATGTGGCGG 1091
135 luGluThrGlnGluLeuArgAlaAlaArgSerArgLeuArgAspPro 151
      :::: |||||
1092 CATGTGCCAGCGCCACGCGAGCTGCAGCATCAACGAGGACA ..... 1132
152 IleLysProGlyMetPheGlyTyArgValProPheAlaLeuProLe 168
1132 ..... 1132
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
      ||| ||| ||| |||||
1133 .... CGGGCTGCCGTGGCCCTTCACTGTAGCCACGAGCTGGGCACAG 1178
185 euLleSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
      :::: ||||| |||||
1179 TTTTGGCATTCAGCATGACGAGCGGCAATGACTGTGAGCCCGTTGGGA 1228
202 GluProPheSerAla ..... AsnGlySerProGlnThrGluLeuPro. 215
      :::: ||||| |||||
1229 AAGCACTTTTCATCATGCTCCACAGCTCTGTACGACGCGCTCCCTC 1278
216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeus 232
      ||| :::: ||||| |||||
1279 ACCTGGTCCGCTGCAGCGCGCATATATACACAGGTTCTCTTGACCGTGG 1328
232 exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
      :::: ||| ||| ||| |||||
1329 GTGGGCGCTGTCCCTGGACGACCCCTCTGATGTAAGCCACGAGTGGCGCTC 1378
249 LeuArgHisHis. .... ProArgAlaGlnAlase 258
      ||| |||
1379 CCTCGTGGCACCTGGGFCCTCTATGATGTAAGCCACGAGTGGCGCTC 1428
258 rGlyThrGluProProSerPro. .... T 266
      :::: ||| |||
1429 CAGTACGGGCGCTACTCTGCTCTGCGAGGACATGGATATGTGTGCCA 1478
266 hrHisSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArg 282
      ||||| :::: |||
1479 CACACTCTGTGCTCTGTGGGA ..... CCACCTGTCT 1510
283 ArgProSer. .... SerGlnGlyTrpAlaSerProGlnValAlaGlyArg. 297
      ||||| ||| ||||| |||||
1511 ACTCCAACTGGATGACCGGTGGACGACCCCGTGTG. .... GGGAGAT 1557
298 ..... ArgProAspProPheProSerValProArgG 308
      :::: |||||
1558 AAGTGGTGTCTCAGTGGGAGTGGCTACCGTGGGCTTCGGGCGCGAGG. 1606
308 lyArgGlyGlnGlnGlyGlnGlyProTrp ..... GlyThrGlyGly 321
      ||||| ||| :::: |||
```

```

610 hrProGluProValProGlnLeuGlnProGluLeuArgValGlu 626
2418 .....
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlu 643
2419 .....ACCATCCA 2426
643 nArgGlnValArgIleProGlnMetProAlaProProHisProArgThrP 660
2427 CAGGGAG .....GCAGGTGCCACGACGAGGTCC 2455
660 roLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676
2456 CG.....CCGCCCTGTCTCTCTGCGATTATGGCCCTGGACCAAGTGC 2499
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693
2500 ACAGTCACCTGGCGAGAGGTGTGCAGAGCAGATGTGTACTGTGTGA 2549
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710
2550 GCGGAGGCGAGG...CCGCTGGAGGAGGACACTGTGACCCCTGGGCC 2596
710 rGProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726
2597 GGCCTGATGACCAACAGAGGAAGTGCAGCAGCAGCCTGCTGCCAGG 2646
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...GI 742
2647 TGGTGGCAGGTGAGTGGCAGGTGTGCTCCAGCTCTCGGGCCTGGGG 2696
742 yThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGly 758
2697 CCTCTCCCGCGGGCGTCTCTGTCATCGCAGCGTGGGCTGGATGAGC 2746
758 lYserSerValProGluArgCysGlyHisLeuProArgProAsnIle 774
2747 AGAGCGCCCTGGAGCCACCGCGCTGTGAACACCTTCCCGCGCCCTACT 2796
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValGlu 789
2797 GAACCCCTTGCACCGCATGATCCCTGTCGCGCCACTGGGCTGTGGG 2846
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSera 806
2847 GAAC...TGGTCTCAGTCTCAGTGCATGTGGGAGGCGCACTCAGCGCC 2893
806 rGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822
2894 GAATGTCTCTGCACCATGACCGGTGTCCCTGTGAGGAGCCGAG 2943
823 CysAlaSerGlyProProGlnProProSerArgGluAlaCysAspMetG1 839
2944 .....CAGCCAGCCAGCGAAGTCACCTGCTCTCTGCC 2975
839 yProCysThr...ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854
2976 ACTCTGTGCGGTGGCCCTGGGCACAC.....TGGGCCCTGAAGGCTCAG 3019
855 SerProGluProProAlaIleSer 862
3020 GCAGCGGCTCTCTCCAGCCACGAGC 3043

```

seq\_name: /SID1/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AD04475

seq\_documentation\_block:

ID\_AD04475 standard; cdna; 5353 BP.

XX

AC AD04475;

XX

DT 04-JUL-2001 (first entry)

```

XX Human 27875 ADAM-TS cDNA, alternative version.
DE
XX Human: ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angiogenesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW cardiant; tumour; thymoma; vasotropic; cytostatic; virucide; ss.
XX Homo sapiens.
OS
XX
XX Location/Qualifiers
FT Key 36..5096
CDS /tag= a
/product= "Human 27875 ADAM-TS protein"
/transl_except= (pos:1821..1823, aa:Xaa)
FT sig_peptide 36..125
FT mat_peptide 126..5093
FT /tag= b
FT /tag= c
FT /product= "Human mature 27875 ADAM-TS protein"
XX WO200131034-A1.
PN
XX 03-MAY-2001.
XX
XX 25-OCT-2000; 2000WO-US29380.
XX
XX 25-OCT-1999; 99US-0426282.
XX (MILL-) MILLENNIUM PHARM INC.
XX Kapeller-Libermann R, White D;
XX WPI: 2001-300513/31.
XX P-PSDB; AAE00913.
XX
XX Novel isolated polypeptide, 27875, a human ADAM-TS (a disintegrin and
PT metalloproteinase) useful for diagnosis and treatment of disorders of
PT bone, lung, heart, skeletal muscle such as osteoporosis, emphysema,
PT angina
XX
XX Claim 7: Page 115-123; 133pp; English.
XX
XX The present sequence is an alternative version of a cDNA encoding
CC 27875 protein, a human ADAM-TS (a disintegrin and metalloproteinase).
CC Metalloproteinase is a proteolytic enzyme involved in protein
CC maturation, protein degradation, tumour growth, metastasis and
CC angiogenesis. Nucleotides encoding 27875, 27875 protein and its
CC antibodies are useful for preventing, diagnosing and treating
CC 27875 metalloproteinase-related disorders. These disorders
CC include, disorders involving the lung such as congenital anomalies,
CC bronchial asthma, Goodpasture's syndrome, pulmonary alveolar proteinosis,
CC disorders involving heart such as heart failure, cardiac hypertrophy,
CC angina pectoris, myocardial infarction, chronic ischaemic heart disease,
CC disorders involving the skeletal muscle include tumours such as
CC rhabdomyosarcoma, disorders involving blood vessel such as hypertension,
CC atherosclerosis, vasculitis associated with other disorders, disorders
CC involving the testis and epididymis such as nonspecific epididymitis and
CC orchitis, gonorrhea, mumps, tuberculosis and syphilis, spermatocytic
CC seminoma, disorders involving the kidney such as membranoproliferative
CC glomerulonephritis, necrotising glomerulonephritis, renal artery
CC stenosis, chronic glomerulonephritis, sickle cell disease nephropathy,
CC urinary tract obstruction, disorders of the bone such as achondroplasia,
CC osteoporosis, Paget's disease, rickets, osteomalacia, osteonecrosis,
CC tuberculous osteomyelitis, osteoma, osteosarcoma, metastatic tumours,
CC and disorders involving thymus such as thymic cysts, thymic hypoplasia,
CC

```

CC thymomas, and Hodgkin's disease. 27875 sequence is used in gene therapy.  
 CC Note: This sequence SEQ.ID.NO.2 is stated as being similar to the  
 CC sequence shown in Fig 1 (See AAD04494). However these sequences differ  
 CC at several locations.

XX

SQ Sequence 5353 BP; 960 A; 1833 C; 1648 G; 911 T; 1 other;

# alignment\_scores:

Quality: 806.50 Length: 959  
 Ratio: 1.816 Gaps: 41  
 Percent Similarity: 46.298 Percent Identity: 28.050

## alignment\_block:

US-10-041-770-2 x AAD04475

Align seg 1/1 to: RAD04475 from: 1 to: 5353

```

15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
   :::::  :::::  :::::  :::::  :::::
631 GTGCTCCAAAGCACCTGTGGAGTGCAAGTATCCAGAGCTGGAGTCTCGA 680
   :::::  :::::  :::::  :::::  :::::
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
   :::::  :::::  :::::  :::::  :::::
681 CGGGACCGTTGGGAGCAGCGGACGAGTGGCGGCGCCACGCGTGAGCG 730
   :::::  :::::  :::::  :::::  :::::
35 hrProThrGluGluGlnGlyProGluGlyValTrpGlyProTrp... 50
   :::::  :::::  :::::  :::::  :::::
731 TCTATACACCGCGTGGTGTGAGCAAGAGAGTGGGTGGAGACCGCTGGTAG 780
   :::::  :::::  :::::  :::::  :::::
50 ..... 50
781 TAGCTATGCCAATAATGGTGGAGTACCAACGACGCGGAGCTTGAGAC 830
   :::::  :::::  :::::  :::::  :::::
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
   :::::  :::::  :::::  :::::  :::::
831 TATGTCTGACCATCATGAACATGGTGGTGGCTTTTCATGACCCAG 880
   :::::  :::::  :::::  :::::  :::::
61 lyValGlyVal.....Gln 65
   :::::  :::::  :::::  :::::  :::::
881 CATTTGGGAACCCCATCCACATACCATTGTGGCGCTGCTGCTGGAAG 930
   :::::  :::::  :::::  :::::  :::::
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80
   :::::  :::::  :::::  :::::  :::::
931 ATGAGGAGGAGGACCTTAAGATCAGCACCATCATGACACACACCTGAAG 980
   :::::  :::::  :::::  :::::  :::::
81 .....SerLeuProLeuP 85
   :::::  :::::  :::::  :::::  :::::
981 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCCATCC 1030
   :::::  :::::  :::::  :::::  :::::
85 roProArgProProArgHisProGluAlaLeuProArgGlyGlnGly 101
   :::::  :::::  :::::  :::::  :::::
1031 CCTGCACCATG..... 1041
102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118
   :::::  :::::  :::::  :::::  :::::
1042 ACACCTGCATCTCTGCTCACCAAAA.....AGGACCTGTGT 1076
   :::::  :::::  :::::  :::::  :::::
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
   :::::  :::::  :::::  :::::  :::::
1077 GCAGCCATGAACCGCCCTGTGAGACCTGGGACTGTCCCATGTGGCGGG 1126
   :::::  :::::  :::::  :::::  :::::
135 luGluThrGlnGluLeuArgAlaAlaArgSerArgLeuArgaspPro 151
   :::::  :::::  :::::  :::::  :::::
1127 CATGTGCCAGCGCACCGCAGCTGCAGCATCAACGAGGACA..... 1167
   :::::  :::::  :::::  :::::  :::::
152 IleLysProGlyMetPheGlyTrpGlyArgValProPheAlaLeuProLe 168
   :::::  :::::  :::::  :::::  :::::
1167 ..... 1167
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
   :::::  :::::  :::::  :::::  :::::
1168 ....CGGGCTGCGCGCTTCACTGTAGCCACGAGCTCGGGCACAG 1213

```

```

185 euLeSerSerArgGlyGluGluAlaLleProSerProThrProArgAla 201
   :::::  :::::  :::::  :::::  :::::
1214 TTTTGGCAATCAGCATGACGGAAGCGCAATGACTGTGAGCCCTGGGA 1263
   :::::  :::::  :::::  :::::  :::::
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
   :::::  :::::  :::::  :::::  :::::
1264 AACGACCTTTTCATCATGTCTCCACAGCTCCTGTACGAGCGCGCTCC 1313
   :::::  :::::  :::::  :::::  :::::
216 ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
   :::::  :::::  :::::  :::::  :::::
1314 ACCTGGTCCCGCTGCAGCGCCAGTATATACACAGGTTCCTTGACCGTG 1363
   :::::  :::::  :::::  :::::  :::::
232 erProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
   :::::  :::::  :::::  :::::  :::::
1364 GTGGGGCTGTGGCTGGAGACCCCTCTCCCAAGGACATTATCGACTCC 1413
   :::::  :::::  :::::  :::::  :::::
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
   :::::  :::::  :::::  :::::  :::::
1414 CCTCGTGCCACCTGGCTGCTCTATGATGTAAAGCCACAGTGGCGCTC 1463
   :::::  :::::  :::::  :::::  :::::
258 rGlyThrGluProProSerPro.....T 266
   :::::  :::::  :::::  :::::  :::::
1464 CAGTACGGGGCTACTCTGCTTCTGCGAGGACATGGATAATGTCTGCCA 1513
   :::::  :::::  :::::  :::::  :::::
266 hrHisSerLeuGlyGlyGlyPheArgAlaSerProGlnProArg 282
   :::::  :::::  :::::  :::::  :::::
1514 CACACTCTGGTGTCTGTGGGA.....CCACCTGTC 1545
   :::::  :::::  :::::  :::::  :::::
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297
   :::::  :::::  :::::  :::::  :::::
1546 ACTCCAGCTGGATGCACCGCTGGACGCCCGGTGTG...GGAGAAT 1592
   :::::  :::::  :::::  :::::  :::::
298 .....ArgProAspProPheProSerValProArgG 308
   :::::  :::::  :::::  :::::  :::::
1593 AAGTGGTGTCTCAGTGGGAGTGGTACCCGTGGGCTTCCGGCCCGAGG. 1641
   :::::  :::::  :::::  :::::  :::::
308 lyArgGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321
   :::::  :::::  :::::  :::::  :::::
1642 .....CCGTGGATGGTGGTGGTGGCTGGCTGA 1668
   :::::  :::::  :::::  :::::  :::::
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338
   :::::  :::::  :::::  :::::  :::::
1669 GCGCCT...GGTCCA..... 1680
   :::::  :::::  :::::  :::::  :::::
338 pleuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355
   :::::  :::::  :::::  :::::  :::::
1691 .....TCTGCTACGGAGCTGTGCGATGGCGGTACA 1711
   :::::  :::::  :::::  :::::  :::::
355 heAlaProSer...Ser.....ProIleProArg..... 363
   :::::  :::::  :::::  :::::  :::::
1712 GAGCGCGAGCGGAGTGCACGCGACCTACGCCCAATACAAAGGCAGAT 1761
   :::::  :::::  :::::  :::::  :::::
364 ..CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379
   :::::  :::::  :::::  :::::  :::::
1762 ACTGTGGTGGTGAAGCGCTCCGCTCTGCAACCTGCAGCGCTGC 1811
   :::::  :::::  :::::  :::::  :::::
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396
   :::::  :::::  :::::  :::::  :::::
1812 CCTGCTGGCCCGCTCTCCGACGCTCCAGTCCAGCCACTTTGACG 1861
   :::::  :::::  :::::  :::::  :::::
396 rGlnGluPheMetGlyGlnLeuTyGlnTrpGluProPheThrGluValG 413
   :::::  :::::  :::::  :::::  :::::
1862 TATGCTCTACAAAGGCCAGCTGCACACATGGTGGTGGCGTG.....GTCA 1905
   :::::  :::::  :::::  :::::  :::::
413 lnGlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429
   :::::  :::::  :::::  :::::  :::::
1906 ATGACGTGAACCCCTGCGAGCTGCAGTCCCGCGCCCGCAATGAGTACTTT 1955
   :::::  :::::  :::::  :::::  :::::
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro... 445
   :::::  :::::  :::::  :::::  :::::
1956 GCCGAGAAGCTGCGGGACGCGGTGCTGATGCGACCCCTCTACCAGGT 2005
   :::::  :::::  :::::  :::::  :::::

```



PS

XX Claim 1b; Page 39-43; 66pp; German.

This invention describes a novel metalloprotease containing a thrombospondin domain (MPTS protein) which is useful to treat aggrecan associated disease including rheumatoid arthritis and osteoarthritis. The products of the invention have osteopathic, antiarthritic, antiproliferative and antirheumatic activity. MPTS may be used to identify a modulator of its activity, e.g., an agonist or antagonist. Such compounds, as well as MPTS itself may be used to treat disease associated with MPTS activity or aggrecan degradation, particularly osteoarthritis, rheumatoid arthritis, psoriasis, spondylitis, sport injury, joint trauma or fibrosis. This sequence encodes the human MPTS-19 metalloprotease described in the method of the invention.

XX

SQ Sequence 5338 BP; 974 A; 1825 C; 1638 G; 901 T; 0 other;

## alignment\_scores:

Quality: 804.50 Length: 959  
Ratio: 1.808 Gaps: 41  
Percent Similarity: 46.403 Percent Identity: 27.946

## alignment\_block:

US-10-041-770-2 x AAH49372 ..

Align seg 1/1 to: AAH49372 from: 1 to: 5338

```

15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
   : : : : : : : : : : : : : : : : : : :
608 GTGCTCCAAAGACCTGTGGAGTGCAAGTGATCCCGAGAGCTGGAGCTCGA 657
   : : : : : : : : : : : : : : : : : : :
24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
   : : : : : : : : : : : : : : : : : : :
658 CGGAGCGTGGGACAGCGGAGCAGTGGCGCGCCAGCGCTGAGCG 707
   : : : : : : : : : : : : : : : : : : :
35 hrProThrGluGlnGlnGlyProGluGlyValTrpGlyProTrp... 50
   : : : : : : : : : : : : : : : : : : :
708 TCTACACCGCGTGGTCCAGCAAGAGAGTGGGTGGAGCCCTGGTGG 757
   : : : : : : : : : : : : : : : : : : :
50 ..... 50

758 TAGCTATCCCAANTGGTGGAGTACCAGGACAGCGGAGCTTGAGAGC 807
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
   : : : : : : : : : : : : : : : : : : :
808 TATGTGCTGACCATCATGAACATGTGTGGCTGGCTGTTTCATGACCCCGAG 857
   : : : : : : : : : : : : : : : : : : :
61 lyValGlyVal.....Gln 65
   : : : : : : : : : : : : : : : : : : :
858 CATTGGGAACCCCATCCACATCACCATTGTGGCGCTGCTCTGCTGGAG 907
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80
   : : : : : : : : : : : : : : : : : : :
908 ATGAGGAGGAGGACCTAAAGATCAGCCACCATGCAGACAACACCCCGAAG 957
81 .....SerLeuProLeuP 85
958 AGCTTCTGCAAGTGGCAGAAAGCATCAACATGAAGGGGAGTGCCTATCC 1007
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
1008 CTGCGACCATG..... 1018
102 ProArgProGlnThrSerProGluThrLeuProLeuTrpArgThrGlnSe 118
1019 ACACGTGCCATCTGCTCCACAGAA.....AGGACCTGTGT 1053
118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
   : : : : : : : : : : : : : : : : : : :
1054 GCAACCATGAACCGCCCTGTGACCCCTGGGACCTGTCCCATGTGGCGGG 1103
135 luGluThrGlnGluLeuArgAlaAlaArgSerArgSerArgLeuArgAspPro 151
   : : : : : : : : : : : : : : : : : : :

```

```

1104 CATGTGCCAGCCGACCGCAGCTGCAGCATCAACAGGAGCA..... 1144
152 IleLysProGlyMetPheGlyTyrglyValPropheAlaLeuProLe 168
1144 ..... 1144
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
   : : : : : : : : : : : : : : : : : : :
1145 ...CGGCGCTGGCGGCTTACATGTAGCCACGAGCTCGGCACAG 1190
185 euLeuSerSerArgGlyGluAlaLeuProSerProThrProArgAla 201
   : : : : : : : : : : : : : : : : : : :
1191 TTTTGGCATTCAGCATGACGAGGAGCAATGACTGTGAGCCCGTTGGGA 1240
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
   : : : : : : : : : : : : : : : : : : :
1241 AACGACCTTTCATCATGCTCCACAGCTCTGTACGACGCGCTCCGCTC 1290
216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
   : : : : : : : : : : : : : : : : : : :
1291 ACTTGGTCCCGCTGCAGCCGCGCATATCACCAGGTTCTTGTGACCGTGG 1340
232 erProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
   : : : : : : : : : : : : : : : : : : :
1341 GTGGGCGCTGTGCTGGAGCAGCCCTCTGCCAAGGACATTATCGACTTC 1390
249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258
   : : : : : : : : : : : : : : : : : : :
1391 CCTCGTCCGACCTGGCGTCTCTATGATGAAGCCACGAGTGGCGCTC 1440
258 rGlyThrGluProProSerPro.....T 266
   : : : : : : : : : : : : : : : : : : :
1441 CAGTACGGGCGCTACTCTGCTCTGCGAGGACATGATAATGTCGCCA 1490
266 hrHisSerLeuGlyGluGlyPheArgAlaSerProGlnProArg 282
   : : : : : : : : : : : : : : : : : : :
1491 CACACTGTGTGCTCTGTGGGA.....CCACCTGTCT 1522
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg.. 297
   : : : : : : : : : : : : : : : : : : :
1523 ACTCCAAGCTGGATGCAGCGCTGCAGCGCACCCGGTGTG...GGAGAAT 1569
298 .....ArgProAspProPheProSerValProArg 308
   : : : : : : : : : : : : : : : : : : :
1570 AAGTGGTGTCTCAGTGGGAGTGCATACCGTGGGCTTCCGCGCCGAGG. 1618
308 lyArgGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321
   : : : : : : : : : : : : : : : : : : :
1619 .....CGTGGATGGTGGCTGTCTGCTGCTGGA 1645
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338
   : : : : : : : : : : : : : : : : : : :
1646 GCGCT...GTCCA..... 1657
338 pLeuProLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355
   : : : : : : : : : : : : : : : : : : :
1658 .....TCTGTCACGGAGCTGTGCATGGCGCTACA 1688
355 heAlaProSer...Ser.....ProIleProArg..... 363
   : : : : : : : : : : : : : : : : : : :
1689 GAGCGCCGAGCGCAGTGCAGCGAGCTACGCCCAATACAAAGGCAGAT 1738
364 ..CysSerGlyCysSerGluGlnLeuArgAlaCysSerGlnAlaProCys 379
   : : : : : : : : : : : : : : : : : : :
1739 ACTGTGTGGTGAGCGCAAGCGCTTCCGCTCTGCAACCTGCAGCGCTGC 1788
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAnSe 396
   : : : : : : : : : : : : : : : : : : :
1789 CTGTGTGGCGCCCTCTCTCCGCCAGCTCCAGTCAGCCACTTTGACGC 1838
396 rGlnGluPheMetGlyGlnLeuTrpGluProPheThrGluValG 413
   : : : : : : : : : : : : : : : : : : :
1839 CATGCTCTACAAGGCGCGCTGCACACATGGGTGCGCGTG.....GTCA 1882

```







354 upheAlaProSer.SerProIleProArg..... 363  
1706 CTTCTAGCGCTCAGTCGACACCCAGCCCAACCATCGGGGGCAAGTAC 1755  
364 CysSerGlyLeuSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380  
1756 TGTCTGGGTGAGAGAGGGCGGACCGCTCTCTCAACAGGATGACTGTCC 1805  
380 oProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397  
1806 CCTTGGCTCCAGGACTTCAGAGAGTGCAGTGTCTGAATTGTACAGCA 1855  
397 InGluPheMetGlyGlnLeuTyrlntPrGluProPheThrGluValGln 413  
1856 TCCTTTTCGTGGGAATCTCAAGTGGAAACGTAC...CGGGGAGGG 1902  
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430  
1903 GCGGTGAAGCCCTGCTCGCTCAGTCCCTAGCGGAAGGCTTCAACTTCA 1952  
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
1953 CAGGAGAGGGCGGACCGTGTGGACGGACACCTGCCGTCCAGACA 2002  
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463  
2003 CGGTGACATTCGTCACTGCGGCAATGAAGACACGTGGGCTGCGACCGA 2052  
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480  
2053 GTCTGGCTCCGACCTCGGGGAGGACAGTCCGAGTGTGGCGGTGA 2102  
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
2103 CGGCAGTGTGCGGACCATCAGGCGCTTTCACCCAGCCCTCACCTG 2152  
497 roLeuGlyTyrlntLysIleLeuTyrlntProAlaGlyAlaLeuArgLeu 513  
2153 GGGCGGTACGAGGATGCTGTGATTCCCAAGGCTCCGTCACATC 2202  
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrlntLeuAlaLeuArgGlyP 530  
2203 TTCATCCAGGATCTGAACCTCTCTCAGTCACTTGGCCCTGAAGGGAGA 2252  
530 oGlyGlyArgSerIleLeuAsnGlyAsnTrpAlaValAspProGlyLys 547  
2253 CCAGGAGTCCCTGCTGTGAGGGGCTGCCCGGACCCCCACCCACC 2302  
547 erTyrlArgAlaGlyThrValPheArgTyrlntAsnArgProProArgGlu 563  
2303 GTCTGCTCTAGTGGGACCACTTTCACTGGCAGACAGGGGCGACCCAG 2352  
564 GluGlyLysGlyGluSerLeuAlaGluGlyProThrThrGlnProVa 580  
2353 GTC.....CAGACCTCGAAGCCTGGACCGCATTAATCATCTCT 2393  
580 laSpValTyrlMetIlePheGlnGluAsnProGlyValPheTyrlnt 597  
2394 CATGCTCATGCTGTGTCGCGGACCGAGCTGCCCTCCCTCCGTACCGCT 2443  
597 yrValIleSerSer.ProProIleLeuGluAsnProThrProGluPr 613  
2444 TC.....AATGCCCATCGCGGTGACTGCTGCCCC...CCTACTCC 2484  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
2485 TGGCATTATGCGCCCTGGACCAAGTGTGCGGCCACAGTGTGCAGCGGTAG 2534  
622 ..... 626  
2535 CCAGGTGACGGCGGTGAGTGCCTCCCAACCAAGCTGGACAGCTCCGCGGTG 2584

626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2585 CCCCCACTACTGTCAGTGCACACCAAGCT.....GCC 2619  
643 InArgGlnValArgIleProGln.MetProAlaProHisProArgTh 659  
2620 AAAGGAGCGCGCTGCACACGAGCCTGCTCCAGAC..... 2661  
659 rProLeuGlySerProAlaAlaTyrlntLysArgValGlyHisSerAlaC 676  
2662 .....TGGGTTGTAGGAAGTGGTCTGCTCT 2686  
676 ySerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysile 692  
2687 GCAGCGCAGCTGCAGTCAGCGCTGGCAGCCCTCGGCTGGTGTGCAG 2736  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706  
2737 CGCGCGTCTCTCGCGGAGGAGAGCGCTGGACGACAGCGCATGCC 2786  
706 aAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProC 723  
2787 GCAG...CCGCGCCACCTGTACTG...GAGGCTGCCACGCGCCACTT 2830  
723 ysProProTyrlntPrGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
2831 GCCTCCGAGTGGCGGCCCTCGACTGCTGTGAGTCACCCCGCAGCTGC 2880  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG 756  
2881 GGGCGGCGCTCCGCCACCGCTGCTTTCAGAGAGC.....GCAGA 2924  
756 yGlyGlySerSerValProGluArgCysGlyHisLeuProArgProA 773  
2925 CCACCGCGCAGCTGCCCGCGGCACTGTACCCCGCGCCCAAGCCAC 2974  
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
2975 CGCGCACCATGCTGCAACTGCCCGCTGCCCGCGCGCTGGGTG 3024  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
3025 GCTGGCGAG...TGGGTGAGTGTCTGCACAGTGCAGCGTGGGSCACG 3071  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
3072 GCAGGCTCGTGTGCTGCACGACGACGCGGCGCAG.....GGTGGC 3115  
821 InGluCysAlaSerGlyProGlnProProSerArgGluAlaCysAsp 837  
3116 ACAGTGCACGAGGCG...CTGGCGCGCCCGCCACGACGACGAGTGTGAG 3162  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT-AAF82157  
seq\_documentation\_block:  
ID AAF82157 standard; cDNA; 3312 BP.  
XX AAF82157;  
AC AAF82157;  
XX 02-JUL-2001 (first entry)  
DT Human ADAM type metal protease MDT2 encoding cDNA SEQ ID NO:11.  
XX Human; a disintegrin and metalloprotease type metal protease; MDT2;  
XX MDT2; ADAM type metal protease; cytosolic; antiarthritic;  
XX cancer; arthritis; arthrosis deformans; ss.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
PH 1..3312  
FT CDS  
FT /\*\*tag= a  
FT /product= "MDT2"

FT /note= "a disintegrin and metalloprotease (ADAM)  
 FT type metal protease"

JP2001008687-A.

16-JAN-2001.

25-JUN-1999; 99JP-0180973.

25-JUN-1999; 99JP-0180973.

(YAMA ) YAMANOUCHI PHARM CO LTD.

WPI: 2001-285362/30.

P-PSDB; AAB74945.

XX New metal protease and metal protease gene, for use as a drug for  
 PT cancers, arthritis and arthrosis deformans

XX Example 2; Page 22-23; 31pp; Japanese.

PS The present sequence encodes a disintegrin and metalloprotease (ADAM)  
 CC type metal protease designated MDS2, isolated from human. MDS proteins  
 CC have cytoskeletal and antiarthritic activities. They can be used as a drug  
 CC for cancers, arthritis and arthrosis deformans.

XX Sequence 3312 BP; 622 A; 1107 C; 1032 G; 551 T; 0 other;

alignment\_scores:

Quality: 796.50 Length: 954  
 Ratio: 1.691 Gaps: 45  
 Percent Similarity: 49.371 Percent Identity: 28.512

alignment\_block:

US-10-041-770-2 x AAF82157 ..

Align seg 1/1 to: AAF82157 from: 1 to: 3312

4 TrpThrGlyArgProTrpLeuTyrLeuLeuLeuLeuLeuLeuLeuProGlu 20  
 |||||:||||||| |||||  
 604 TGGAAAGGGGGCCATGG...TGGCTGGGACCTTGAAGCCACCGCCTGC 650  
 20 nLeucysLeuaspGlnGluValLeuSerGlyHisLeuGlnThrProT 37  
 |||||:||||||| |||||  
 651 CAGGCCCTCGGGGAATGAA.....A 670  
 37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46  
 |||||:||||||| |||||  
 671 CAGAGCGTGGCCAGCCAGCCCTGAAGCGATCGGTGAGCGAGAGCGCTAC 720  
 46 lTrpGlyProTrp..... 50  
 |||||:||||||| |||||  
 721 GTGGAGACCCCTGGTGGTGACAAAGATGATGGTGGCTATCAGGGCG 770  
 51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66  
 :|||:||||||| |||||  
 771 CCGGGATGTGGAGCAGTATGCTCTGCCCATCATGAACATTTGTCACAAAC 820  
 67 ArgSerArgThr..... 70  
 |||||:||||||| |||||  
 821 TTTTCAGGACTCGAGTGTGGGAAGCACCGTTAACATCTCTGTAACCTCGC 870  
 71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85  
 |||||:||||||| |||||  
 871 CTCATCTCTGCTCAGGAGGACCGCCACTCTGGAGATCA..... 910  
 85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 |||||:||||||| |||||  
 911 ..CCCAACATGCCGGGAAGTCCCTAGACAGCTTCTTAAGTGGCAGAAAT 958  
 102 ProArgProGlnThrSerProGlnThrLeuProLeuTyrArgThr..... 116  
 |||||:||||||| |||||

959 CCATCGTGAACACACAGCGCCATGGCAATGCCAATTCACAGAGAACGGTGTG 1008  
 117 .....GlnSerArgGlyArgGlyGlyP 124  
 1009 GCTAACCATGACACACAGCAGTGTCTCATCACACGCTATGACATCTGCATCTA 1058  
 124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluIle 140  
 :|||:||||||| |||||  
 1059 CAAGAAACAAACCTGGCGCACACTAGGCTGGCCCGGTGGCGGAATGT 1108  
 141 ArgAlaAlaArgArg.....SerArgLeuArgAspProIleLeuProGlu 155  
 :|||:||||||| |||||  
 1109 GTAGCGCGAGAGAGAGCTGACGCTCAATGAGACATTTGGCTGGCCACA 1158  
 155 YMetPheGlyTyrGlyArgValProPheAlaLeuProLeu.HisArgAsn 171  
 |||||:||||||| |||||  
 1159 G.....CGTTCACCATTTGCCACGAGAT 1181  
 172 ArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSerSe 188  
 |||||:||||||| |||||  
 1182 CGGGCACACATTCGGCATGAACA.....TGACG 1210  
 188 rATGlyGluGluAlaIleProSerProThrProArgAlaGluProPheS 205  
 |||||:||||||| |||||  
 1211 CGTGGGAACACCTGTGG.....GGCCGCTGTGAGGACCCAGCC 1251  
 205 erAlaAsnGlySerProGln...ThrGluLeuProProThrGluLeuSer 220  
 :|||:||||||| |||||  
 1252 AAGCTCATGCTGCCACATTAACCATGAAGACCAACCCATTCGTGTGGTC 1301  
 221 ValHisThrPro.....SerProGlnAlaGluProLeuSerProGluTh 235  
 |||||:||||||| |||||  
 1302 ATCTGCGCGCTGACTACATCACCGCTTCTAGACTGGCGCTGGGC 1351  
 235 rAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgHisH 252  
 :|||:||||||| |||||  
 1352 TCTGCTGAACACCGCGCCCGCCAGACGACTTTGT..... 1388  
 252 isProArgAlaGlnAlaSerGlyThrGluPro..... 262  
 |||||:||||||| |||||  
 1389 .....GTACCCGACAGTGGCACCGGGCCAGCCCTACGATCAGATGA 1430  
 263 ProSerProThrHisSerLeuGlyGlu.....G 172  
 |||||:||||||| |||||  
 1431 GCAATGCCGCTTTCAGCATGGAGTCAATCGCTAGTGTAAATACGGG 1480  
 272 yGlyPhePheArgAla.....SerProGlnPro..... 281  
 |||||:||||||| |||||  
 1481 AGGTCTGACGAGCTGTGTGTGTGAGCAAGACCAACCGGTGCATCAC 1530  
 282 ..ArgArgProSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArg 297  
 :|||:||||||| |||||  
 1531 AACAGCATCCGCGCGCGAGGGCAGCTGTGCCAGACGACACCATCGA 1580  
 298 ArgPro.....AspProPheProSerValProAr 307  
 :|||:||||||| |||||  
 1581 CAAGGGTGTGTGTACAACGGGTCTGTGCCCTTTGGTTCGCGCCAG 1630  
 307 gGlyArgGlyGlnGlnGlyGlnGlyProTrpGlyThrGlyGlyThrProH 324  
 |||||:||||||| |||||  
 1631 AGGTGTGGAGCG.....AGCTGGGGCGGTGGACT 1662  
 324 isGlyProArgLeuGluProAspProGlnHisProGlyAlaTrpLeuPro 340  
 |||||:||||||| |||||  
 1663 CCATGGGCGCACTGCAGCGGAC..... 1685  
 341 LeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSer.....Le 354  
 |||||:||||||| |||||  
 1886 .....CTGTGGCGGGCGGTGTCT 1705  
 354 uPheAlaProSer\_SerProfileProArg..... 363  
 |||||:||||||| |||||  
 1706 CTTCTAGCGCTCACTGGCAGACAGCCCGCCAGCAACCATCGGGGCAAGTAC 1755

```
364 CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1756 TGCTGGGTGAGAGAGCGCGCACCGCTCTCTGCAACAGGATGACTGCC 1805
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 pProGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1806 CCCTGGCTCCAGGACTTCAGAGAGTGCAGTGTCTGAATTTGACAGCA 1855
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 InGluPheMetGlyGlnLeuTyrGlnTyrGluProPheThrGluValGln 413
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1856 TCCCTTCCGTGGGAATCTACAGTGGAAACGTAC...CGGGAGGG 1902
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTy 430
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1903 GCGGTGAAGGCTGCTCGCTACGCTACGCTAGCGGAAGGCTTCAACTTCTA 1952
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1953 CACGAGAGGGGGCGGCGCGTGGTGGAGCGGACACCTGCGGTCAGACA 2002
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2003 CGGTGGACATTTGCTAGTCAAGTGGCGAATCAACAGCAGCGTGGCTGCACCGA 2052
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 lleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyLys 480
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2053 GTCTGGGCTCCGACTCGCGGAGGACAAAGTCCGAGTGTGGGGTGA 2102
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyLyp 497
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2103 CGCGAGTGCCTCGAGACCATCGAGGGGCTCTTCAGCCGACGCTCACCTG 2152
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
497 roLeuGlyTyrGlnLysIleLeuTyrIleProAlaGlyAlaLeuArgLeu 513
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2153 GGGCGGGTACGAGGATGCTGCTGGATCCCAAGGTCCTCGTCCACATC 2202
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyPr 530
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2203 TTCATCCAGATCTCAACCTCTCTCACTCACTTGGCCCTGAAGGGAGA 2252
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
530 oGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProProGlyS 547
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2253 CCAGGAGTCCCTGCTGCTGAGGGGCTGCCGAGCCGCCGCCGCCACC 2302
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
547 eTyrArgAlaGlyThrValPheArgTyrAsnArgProProArgGlu 563
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2303 GTCTGCTCTAGTGGGACCACTTTCAACTCGCACAGGGGCGCACAGCAG 2352
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
564 GluGlyLysGlyGluSerLeuAlaGlyProThrThrGlnProVa 580
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2353 GTC.....CAGAGCTCGAAGCCCTGGGACCGATTATGCACTCT 2393
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
580 lAspValTyrMetIlePheGlnGluAsnProGlyValPheTyrGlnT 597
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2394 CATCGTCATGGTGGTGGCCGCGAGCGAGCTGCTGCCCTCCGCTACCGCT 2443
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
597 yValIleSerSer..ProProIleLeuGluAsnProThrProGluPr 613
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2444 TC.....AATGCCCCCATCGCCGCTGATCGTGCCTCC...CCTACTCC 2484
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
613 o.....ProValProGlnLeuGlnProGlu..... 621
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2485 TGGCACTATGCGCCCTGGACCAAGTCTCGGCCGACGTCGAGGGGAG 2534
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
622 .....Ile.LeuArgValol 626
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2535 CCAGGTGACGGGGTGGAGTGGCGCAACAGCTGGGACAGCTCCGGGTG 2584
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2585 CCCCCCAGTACTGCACTGCCAGCAGCAAGCT.....GCC 2619
```

```
643 InArgGlnValArgIleProGln.MetProAlaProProHisProArgTh 659
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2620 AAAAGCAGCGCGCTGCAACACGAGCGCTTGCCCTCCAGAC..... 2661
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
659 rProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlac 676
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2662 .....TGGGTGTAGGAACTGGTCTGCTCT 2686
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
676 ySSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIle 692
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2687 GCAGCGCAGCTCGATGCGAGTGGCGCGCGCGCTCGGTGCGTGCAG 2736
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2737 CGCGCGTCTCTCGCGGAGAGAGCGCTGACGACAGCGCATGCC 2786
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
706 aAlaGlyAlaArgProAlaSerProGluProCysHisGlyThrProC 723
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2787 GCAG...CCGCGGCCACTGTACTG...GAGGCTGCCAGCGGCCACTT 2830
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
723 ySProProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2831 CCCTTCGGAGTGGCGGCCCTCGACTGCTGAGTGCACCCCGCAGCTGC 2880
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly 756
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2881 GGGCGGGGCTCCGCCACCGCGTGGTCTTTGCAAGAGC.....GCAG 2924
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
756 yGlyGlySerSerValProProGluArgCysGlyHisLeuProArgProA 773
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2925 CCACCGCGCCAGCTGCCCGCGGCACTGCTACCCCGCGCCAGCCAC 2974
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2975 CGGCACCATCGGCTGCAACTTGGCGCGCTGCCCGCGCGCTGGGTG 3024
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3025 GCTGGCGAG...TGGGTGAGTGTCTGACAGTGGCGCTCGGGCAGCG 3071
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspLysValSerGlu 821
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3072 GCAGCGCTCGGTGGCTGCACGACCCACAGGGGCGAG.....GCGTCCG 3115
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
821 InGluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3116 ACAGTGCACGAGGCGC...CTGGCGCGGCCACCGACGACGAGTGTGAG 3162
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAH43769
seq_documentation_block:
ID AAH43769 standard; DNA; 3776 BP.
XX
AC AAH43769;
XX
DT 30-JAN-2002 (first entry)
XX
DE ADAMTS-E DNA.
XX
KW ADAMTS-E; disintegrin; metalloprotease; thrombospondin domain; stroke;
KW ADAMTS-E2; splice variant; osteoarthritis; rheumatoid arthritis; pain;
KW inflammatory bowel disease; Crohn's disease; emphysema; infertility;
KW acute respiratory distress syndrome; spinal cord injury; head trauma;
KW chronic obstructive pulmonary disease; Alzheimer's disease; asthma;
KW organ transplant toxicity; rejection; cachexia; allergy; cancer;
KW tissue ulceration; restenosis; periodontal disease; Parkinson's disease;
KW epidermolysis bullosa; osteoporosis; atherosclerosis; aortic aneurysm;
KW congestive heart failure; myocardial infarction; cerebral ischaemia;
KW neurodegenerative disease; Huntington's disease; peripheral neuropathy;
KW cerebral amyloid angiopathy; amyotrophic lateral sclerosis; migraine;
KW multiple sclerosis; ocular angiogenesis; corneal injury; depression;
KW macular degeneration; abnormal wound healing; burns; diabetic shock; ds.
XX
```

OS	Homo sapiens.
XX	
FFH	Key Location/Qualifiers
CDS	139..3453
FT	/tag= a
FT	/product= "ADAMTS-E"
XX	
PN	EP1149903-A1.
PD	31-OCT-2001.
XX	
XX	26-APR-2001; 2001EP-0303818.
XX	
PR	26-APR-2000; 2000US-199924P.
PPT	(PFZ ) PFIZER PROD INC.
PFA	
XX	Bucklinder L, Mitchell PG, Walsh RT, Wachtmann TS;
PPI	WPI; 2002-019321/03.
DR	P-PADB; AAB47719.
XX	
XX	Novel polynucleotide encoding ADAMTS-E, a disintegrin and metalloproteinase containing thrombospondin domain, and ADAMTS-E2 polypeptide useful for treating cerebral ischemia, osteoporosis, cachexia, allergies and asthma -
PT	
PT	Claim 1; Page 16-17; 48pp; English.
PS	This sequence encodes ADAMTS-E which is a disintegrin and metalloprotease containing a thrombospondin domain. ADAMTS-E2 is a splice variant of ADAMTS-E. The ADAMTS-E DNA is useful for diagnosing a disease or susceptibility to a disease in a subject relative to expression or activity of ADAMTS-E in a subject. ADAMTS-E and the splice variant may be used in the treatment of osteoarthritis and rheumatoid arthritis, inflammatory bowel disease, Crohn's disease, emphysema, acute respiratory distress syndrome, asthma, chronic obstructive pulmonary disease, Alzheimer's disease, organ transplant toxicity and rejection, cachexia, allergy, cancer, tissue ulcerations, stenosis, periodontal disease, epidermolysis bullosa, osteoporosis, loosening of artificial joint implants, atherosclerosis, aortic aneurysm, congestive heart failure, myocardial infarction, stroke, cerebral ischemia, head trauma, spinal cord injury, neurodegenerative diseases, autoimmune disorders, Huntington's disease, Parkinson's disease, migraine, depression, peripheral neuropathy, pain, cerebral disease, amyloid angiopathy, nootropic, or cognition enhancement, amyotrophic lateral sclerosis, multiple sclerosis, ocular angiogenesis, corneal injury, macular degeneration, abnormal wound healing, burns, infertility or diabetic shock.
XX	
SQ	Sequence 3776 BP; 710 A; 1253 C; 1202 G; 611 T; 0 other;
alignment_scores:	
Quality:	796.00 Length: 954
Ratio:	1.697 Gaps: 44
Percent Similarity:	49.161 Percent Identity: 28.407
alignment_block:	
US-10-041-770-2 x AAH43769 ..	
Align seg 1/1 to: AAH43769 from: 1 to: 3776	
4 TrpThrGlyArgProTrpLeuLeuLeuLeuLeuSerLeuProGl 20 	
742 TGGAAGGGCGGCCATGG...TGCGTGGACCTTGAAGCCACCGCTGC 788 	
20 nLeuCysLeuaspGlnGluValLeuSerGlyHisSerLeuGlnThrProt 37 	
789 CAGGCCCTGGGGAATGAA.....GlyProGlu.GlyVa 46 	

```

alignment_scores:
  Quality: 796.00      Length: 954
  Ratio: 1.697         Gaps: 44
  Percent Similarity: 49.161  Percent Identity: 28.407

alignment_block:
  US-10-041-770-2 x AAH43769 ..

Align seg 1/1 to: AAH43769 from: 1 to: 3776

      4 TrpThrGlyArgProTprLeuTyLeuLeuLeuLeuSerLeuProG1 20
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      742 TGGAAAGGGCGCCATGG...TGCGTCGGACCTTGAAGCCACCGCTGC 788

      20 nLeuCysLeuAspGlnValLeuSerGlyHisSerLeuGlnThrProT 37
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      789 CAGGCCCTGGGGAATGAA.....A 808

      37 hrGluGluGlyGln.....GlyProGluGlyVa 46
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

564 GluGlyLysGlySerLeuSerAlaGluGlyProThrThrGlnProVa 580  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2491 GTC.....CAGAGCCCTCGAAGCCCTGGACCGGATTAATGCATCTCT 2531  
  
580 lAspValTyrMetIlePheGlnGluAsnProGlyValPheTyrGlnT 597  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2532 CATCGTCATGGTGCTGCCCGGACCAGGTGCTGCCCTCGCTACCGCT 2581  
  
597 yrValIleSerSer.ProProProlIleLeuGluAsnProThrProGluPr 613  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2582 TC.....AATGCCCATCGCCCGCTGACTCGCTGCCCTGCCCTCC 2622  
  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2623 TGGCACTATGCCCTCGCACCAAGTCTCGGCCAGTGTGCAGCGGTAG 2672  
  
622 .....Ile.LeuArgValGI 626  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2673 CCAGGTGCAGCGCGGTGGAGTGC CGCAACAGCTGGACAGCTCCGCGGTG 2722  
  
626 uProProLeuAlaProAlaProAlaProAlaArgThrProGlyThrLeug 643  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2723 CCCCCCACTACTGCAGTGCACACAGCAAGCT.....GCC 2757  
  
643 lnArgGlnValIleProGln.MetProAlaProProHisProArgTh 659  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2758 AAAAGGCAGCGCGCTGCAACACAGGAGCTTGCCCTCCAGAC..... 2799  
  
659 rProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlac 676  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2800 .....TGGTTGTAGGAACCTGGCTGCTCT 2824  
  
676 ysSerAlaSerCysGlyValrParpGProIlePheLeuCysIle 692  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2825 GCAGCCGAGCTGCATGCAGCGCTGCAGCGCTCGCTGCTGCTGCCAG 2874  
  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysal 706  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2875 CCGCGCGCTCTCGCGGAGGAGGAGCGCTGCAGACAGCATGCGCC 2924  
  
706 alaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProc 723  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2925 GGAG...CGCGGCCACCTGTACTG...GAGGCTGCCACGCGCCCACIT 2968  
  
723 yspProTyrrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
2969 GCCTCCGGAGTGGCGGCCCTCGACTGTGTGAGTGCACCCCGCAGCTGC 3018  
  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGl 756  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3019 GGCGCGGCTCCGCCACCGGTGGTCTTTGCAAGAGC.....GCAGA 3062  
  
756 yGlyGlySerSerValProProGluArgCysGlyHisLeuProArgProA 773  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3063 CCACCGCGCACGCTGCCCGCGCGGCGCACTGCTCACCGCGCGCCAAAGCCAC 3112  
  
773 snileThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3113 CGGCCACCATGCTGCACACTTGGCCCTGCCCGCGCGCGCTGGGTG 3162  
  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3163 GCTGGCGAG...TGGGGTGTGCTGTCTGCACAGTGCAGCGCTGGCGGACGC 3209  
  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3210 GCAGCGCTGGTGGCTGCACCGACCCACAGCGGCCAG.....GCGTCGC 3253  
  
821 lnGluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837  
:::|||||:::|||||:::|||||:::|||||:::|||||:::  
3254 ACGAGTGCAGGAGCGCTCGCGCGCGCGGCTACACGACGACGCTGGAG 3303  
  
ser name: /STD1/accdData/hold-geneseq/geneseq-emb1/NA2001A.DAT:

seq name: /SIPSI/qcadata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF63445



1323 CCCGACAGTGCACCGGGCCCAAGCCTACGATGCAGATGAGCAATGCCGCT 1372  
266 rHisLeuGlyGlu.....GlyGlyPheA 276  
1373 TTCAGCATGGAGTCAATCGCTCAGTGAATAACAGGGGAGGTCTGCAGC 1422  
276 rGAla.....SerProGlnPro.....ArgArgPro 284  
1423 GAGCTGTGGTGTCTGAGCAAGAGAACCGGTGCATCACCAACAGCATCCC 1472  
285 SerSerGlnGlyTTPAlaSerProGlnValAlaGlyArgArgPro..... 299  
1473 GCCCGCCAGGGCAGCTGTGCCAGACCCACACCATCGACAAAGGGTGGT 1522  
300 .....AspProPheProSerValProArgGlyArg 311  
1523 GCTACAAACGGGCTGTGTCCCTTTGGTGTGCGCCCGAGAGGGTGGAC 1572  
311 lnglnGlnGlyProTrpGlyThrGlyThrProHisGlyProArg 327  
1573 GG.....AGCCTGGGGCGGTGGACTCCATGGGGGA 1604  
328 LeuGluProAspProGlnHisProGlyAlaTrpLeu.ProLeuLeuSera 344  
1605 CTCGACGGGAC.....CTGTGGCGCGCGCTGCTCCTCT 1639  
344 snGlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSerPro 360  
1640 CTAGTCGTCTACTGC.....GACAGCCCC 1662  
361 lleProArg.....CysSerGlyGluSerGluGlnLeuArg 372  
1663 AGGCCAACCATCGGGGCAAGTACTGTGGGTGAGAGAGCGGCACCG 1712  
372 galaCysSerGlnAlaProCysProGluGlnProAspProArgAlaAl 389  
1713 CTTCTGCAACAGGATGACTGTCCCTCGCTCCAGGACITTCAGAGAG 1762  
389 euGlnCysAlaAlaPheAnSerGlnGluPheMetGlyGlnLeuTrpGln 405  
1763 TGCAGTGTCTGAATTGACAGCATCCCTTTCCGTGGGAAATTCACAG 1812  
406 TrpGluProPheThrCluValGlnGlySerGlnArgCysGluLeuAsnCy 422  
1813 TGGAAACCATAC...CGGGAGGGGGGTGAAGGCTGCTCCTCAGCAG 1859  
422 sArgProArgGlyPheArgPheTrpValArgHisThrGluLysValGlnA 439  
1860 CCTAGCGGAAGCTTCACTTCTACAGGAGAGGGCGGCGAGCGGTGGT 1909  
439 spGlyThrLeuGlyGlnProGlyAlaProAspIleCysValAlaGlyArg 455  
1910 ACGGGACACCTCGCCTCCAGACAGGTTGGACATTGGCTCAGTGGCGAA 1959  
456 CysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgProAs 472  
1960 TGCACACAGTGGCTGCGACCGAGTCCCTGGGCTCGGACCTCGGGAGGA 2009  
472 pGlyCysGlyValCysGlyCylAspAspSerThrCysArgLeuValSerG 489  
2010 CAAGTCCGAGTGTGTGGGGTGCAGCGAGTCCCTCGGAGACCATCGAG 2059  
489 lyaAsnLeuThrAspArgGlyGlyProLeuGlyTrpGlnLysIleLeuTrp 505  
2060 GCCTCTTACCCAGCCTCACTCGGGCGGGGTACAGGATGCTCGTCTGG 2109  
506 ileProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSe 522  
2110 ATTCCCAAAGGCTCCGTCACATCTTCAATCCAGGATCTGAACCTCTCT 2159  
522 rAsnTyrLeuAlaLeuArgGlyProGlyArgSerIleIleAsnGlyA 539  
2160 CAGTCACTTGGCCCTGAAGGAGACAGGAGTCCCTGCTGCTCGAGGGGC 2209  
539 snTrpAlaValasPProProGlySerTyrArgAlaGlyGlyThrValPhe 555  
2210 TGCTGGAGACCCCGCCAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2259  
556 ArgTyrAsnArgProProArgGluGluGlyLysGlyGlySerLeuSera 572  
2260 CAAGTGGAGAGGGCCAGACAGGTC.....CAGAGCTCGAAGC 2300  
572 aGluGlyProThrThrGlnProValaspValTyrMetIlePheGlnGlu 589  
2301 CTTGGGACCGATTAAATGCAATCTCTCATCGTCAATGCTGCTGGCCGAGCG 2350  
589 luAsnProGlyValPheTyrGlnTyrValIleSerSer.ProProProI 605  
2351 AGCTGCTGCTCCCTCGCTACCGCTTC.....AATGCCCATCGCCCT 2394  
605 eLeuGluAsnProThrProGluPro.....ProValProGlnL 618  
2395 GACTCGCTGCCCC...CCTACTCTGCTGCACTATGCGCCTTGGACCAAGTG 2441  
618 euGlnProGlu..... 621  
2442 CTCGCCCATGTGTCAGCGCGTAGCCAGGTGCAGGGGTGGAGTGCAGCA 2491  
622 .....Ile.LeuArgValGluProProLeuAlaProAlaProArgP 635  
2492 ACCAGCTGGACAGCTCGCGGTGCGCCCGCTACTGCTGCTGCTGCTGCTG 2541  
635 roAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln.Me 651  
2542 AAGCT.....GCCCAAAGGACGCGCGCTGCACACAGCA 2576  
651 tProAlaProProHisProArgThrProLeuGlySerProAlaAlaTyrT 668  
2577 GCCTTGCCCTCCAGAC..... 2593  
668 rLysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyVal 684  
2594 GGTGTGAGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2643  
685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGlu..... 699  
2644 CGCAGTACGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2693  
700 .....LeuAspGluArgSerCysAlaAlaGlyAlaArgProAlaSerP 715  
2694 GCGCTGGACAGACAGCGATGCCCGCAG...CCGCGCCACCTGTACTG. 2739  
715 roGluProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGlu 731  
2740 ..GAGGCTGCGCCAGCGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2787  
732 TrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLe 748  
2788 TGGTGTGAGTGTACCAAGCTGTGGGCTGCTGCTGCTGCTGCTGCTGCTG 2837  
748 uGlnCysArgGlnGluPheGlyGlyGlySerValProProGluA 765  
2838 CTTTGTGAAGT.....GCAGATCAACGATCTACTGCTGCTGCTGCTGCTG 2881  
765 rGlyCysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArg 781  
2882 ACTGCTTCTCTGCGACCAAGCCACCATCTACTATGCGATGTAACCTGCGC 2931  
782 LeuCys.....GlyHisTrpGluValGlySerProTrpSerGlnCysSe 796  
2932 CGCTGCCCTCTGCCCGTGG...GTGACCAAGTGTGGGTGAGTGTCTC 2978  
796 rValArgCysGlyArgGlyGlnArgSerArgGlnValArgCysValGlyA 813  
2979 CACACAGTGTGGCTCGGCGCAGCAGCAGCAGTGGCTGCACCGCC 3028

CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2964 BP; 705 A; 862 C; 796 G; 601 T; 0 other;

alignment\_scores:  
Quality: 789.50 Length: 366  
Ratio: 3.290 Gaps: 6  
Percent Similarity: 65.574 Percent Identity: 42.350

alignment\_block:  
US-10-041-770-2 x AAH16636 ..

Align seg 1/1 to: AAH16636 from: 1 to: 2964

523 AsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleIleAsnGlyAs 539  
|||||  
3 AACTATTGGCCCTGAGAGTCGTTCTGGACGCTCCATCATCAATGGAA 52  
539 nTrpAlaValAspProGlySerTyrArgAlaGlyGlyThrValPheA 556  
|||||  
53 CTGGCAATTGATGACACAGAAATACGAGGCGGAGGACCATGTTC 102  
556 rGlyAsnArgProProArg...GluGluGlyLysGlyGlySerLeuSer 571  
|||||  
103 CCTACAGGCTCCAAATGAGATTTCGACACTGCGGAGAGTCTCTTTG 152  
572 AlaGluGlyProThrThrGlnProValAspValTyrMetIlePheGlnG 588  
|||||  
153 GCGGAAGGTCCACCAACGAGATCTTGGATGCTACATGATACACACGA 202  
588 uGluAsnProGlyValPheTyrGlnTyrValIleSerSerProProI 605  
: |||||  
203 GCCAAACCCAGCGCTGACATCGAGTACGTGATGATGGGACCAACGCA 252  
605 leLeuGluAsnProThrProGluProProValProGlnLeuGlnPro... 620  
|| |||||  
253 TC.....AGCCCCAGGTGCCA...CCCCACAGGAGACACGGG 287  
620 .....  
288 GAACCTTCAATGGCCAGATGTTGACAGAGCCAGGAGGAGGAGGG 337  
621 .....GluIleLeuArgValGluProP 628  
338 AGNACAGAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 387  
628 roLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnArg 544  
|| |||||  
388 CTGAGATGTTACCTCAGAAATCGGCACACCTTCCACGACGATCCCA 437  
645 GluValArgIleProGlnMetPro.....AlaProProHisProAr 658  
: |||||  
438 GACAGATTTCTCCCTCCACCGGACACTGTTGTCACCCACGACCGCA 487  
658 gThrProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerA 675  
|| |||||  
488 GCCCCACGCGCGCCGCGGATCACAACCTGGAAGCAGCTTGGGACACAG 537  
675 lacysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCys 691  
|||||  
538 AATGTTCCAGCAGCTGTGGGAAGGATCGCAGTACCTATTTCCTCTGT 587

813 snAsnGlyAspGluValSerGluGlnGluCysAlaSerGlyProGln 829  
: |||||  
3029 ACACCGGCAG.....CATCTCGAGAGTGCACF..... 3057  
830 ProProSerArgGluAla.CysAspMetGlyProCysThr..... 842  
|||||  
3058 .....GAGCCTTGGCGCCATCCACCATGACGACGATGTGAGGC 3095  
842 .....  
3096 CAAGTGTGACAGTGTGGTGGCGCTGGAGATGGCCGACGAGATCAAGG 3145  
843 .....ThrAlaTrpPheHisSerAspTrpSerSerLysValSer..... 855  
|| |||||  
3146 ATGTGAACAAGTGTGCTTACTGCCCTTGGTCTCAATTTTCAGTTCTGT 3195  
856 ProGluProProAlaIleSerCys 863  
|||||  
3196 AGCGAGCGCTACTTCCGCGAGAIGT 3220

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH16636

seq\_documentation\_block:

ID AAH16636 standard; cDNA; 2964 BP.

XX AC AAH16636;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:15752.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602  
XX PT full-length cDNAs defined in the specification, and for the detection  
XX PT and/or diagnosis of the abnormality of the proteins encoded by the  
XX PT full-length cDNAs -

XX PS Claim 8; SEQ ID 15752; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602  
XX CC full-length cDNAs defined in the specification. Where a primer set  
XX CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
XX CC to the complementary strand of a polynucleotide which comprises one of  
XX CC the 5602 nucleotide sequences defined in the specification, where the  
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX CC of an oligonucleotide comprising a sequence complementary to the  
XX CC complementary strand of a polynucleotide which comprises a 5'-end  
XX CC sequence and an oligonucleotide comprising a sequence complementary to a  
XX CC polynucleotide which comprises a 3'-end sequence, where the  
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of





```

116 ..... 116
943 GCTAACATGACACAGCAGTCTCATCACAGCGTATGACATCTGCATCTA 992
117 .....GlnSer ArgGlyArgGlyProLeuArgGlyProAl 129
118 ..... 118
993 CAAGAACAAACCTCGCGCACACTAGCCTGCGCGGTGGCGGAATGTG 1042
129 aSerHisLeuGlyArgGluGluThr.....GlnGluLeuArgA 142
130 ..... 130
1043 TGAG.....CGCGAGAAAGCTGACGCGTCAATGAGACATTTGGCT 1083
142 laAlaArgSerArgLeuArgAspProIleLysProGlyMetPheGly 158
143 ..... 143
1084 GCCAACCGGTTC..... 1097
159 TyrGlyArgValProPheAlaLeuProLeuHisArgAsnArgA-GHisPr 175
160 ..... 160
1098 .....CCATTGCCAGAGATCGGCACACAT 1123
175 aArgSerProArgSerGluLeuSerLeuIleSerSerArgGlyGluG 192
176 ..... 176
1124 TGGGCATGAACCA.....TGACGGCGGTGGGAAC 1152
192 luAlaIleProSerProThrProArgAlaGluProPheSerAlaasnGly 208
193 ..... 193
1153 AGCTGTGG.....GCCCGTGTGCAGGACCCAGCCAGCTCATGGC 1193
209 SerProGln.....ThrGluLeuProProThrGluLeuSerValHisThrPr 224
210 ..... 210
1194 TGCCACATATCAATGAAGAACCAACCTCGTGGTCACTCTGCAACC 1243
224 o.....SerProGlnAlaGluProLeuSerProGluThrAlaGlnThrG 239
225 ..... 225
1244 GTGACTACATCACAGCTTCTAGACTCGGGCTGGGCTCTGCCTGAAC 1293
239 luValAlaProArgThrArgProAlaIleProLeuArgHisProArgAla 255
240 ..... 240
1294 AACCGCGCCCGCAGACAGGACTTTGT.....GTA 1322
256 GlnAlaSerGlyThrGluPro.....ProSerProTh 266
267 ..... 267
1323 CCGCAGATGGCAGCGGCGCAAGCTACGATGACAGCAATGCCGCT 1372
266 rHisSerLeuGlyGlu.....GlyGlyPhePheA 276
277 ..... 277
1373 TTCAGCATGGAGTCAATCGCTCGTCAAGTAAATACGGGAGGTGTCAGC 1422
276 rGala.....SerProGlnPro.....ArgArgPro 284
285 ..... 285
1423 GAGCTGTGTGTCTGACGAAGACACCGGTGTCATCACCAACAGCATCC 1472
285 SerSerGlnGlyTrpAlaSerProGlnValAlaGlyArgArgPro..... 299
1473 GGCGCGGAGGCGCTGTGCGCAGACACACCATCGACAGGGGGTGGT 1522
300 ..... 300
1523 GCTACAAACGGTCTGTGTCCCTTGGTGGCGCCAGAGGGTGTGGAC 1572
311 lncLysGlnGlyProTrpGlyThrGlyThrProHisGlyProArg 327
312 ..... 312
1573 GG.....AGCCTGGGGCGGTGGACTCCATCGGGGGA 1604
328 LeuGluProAspProGlnHisProGlyAlaTrpLeu.ProLeuLeuSerA 344
329 ..... 329
1605 CTGACCGCGGAC.....CHGTGGCGGGCGGTGCTCCTCT 1639
344 snGlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSerPro 360
361 ..... 361
1640 CTAGTGGTCACTGC.....GACAGCCCC 1662
361 IleProArg.....CysSerGlyGluSerGluGlnLeuAr 372

```

```

1663 AGCCCAACCATCGGGGCAAGTACTCTCTGGGTGAGAGAAGCGCGCACCG 1712
372 gaAlaCysSerGlnAlaProCysProProGluGlnProAspProArgAla 389
389 ..... 389
1713 CTCTCTGCAACAGCATGACTGTCCCTGCTCCAGGACTTCAGAGAAG 1762
389 euGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTrpGln 405
390 ..... 390
1763 TGCAAGTTCTGAATTGACAGCATCCCTTCGTTGGGAATTCACAG 1812
406 TrpGluProPheThrGluValGlnGlySerGlnArgCysGlnLeuasnCy 422
422 ..... 422
1813 TGGAAAACGTAC...CGGGAGGGGGCGGTGAGGGCTGCTCGTCCAGAG 1859
422 sArgProArgGlyPheArgPheTyrValArgHisThrGluLysValGlnA 439
439 ..... 439
1860 CCTAGCGGAAGGCTTCAACTTCTACCGAGAGGGGCGGCGCGGTGGTG 1909
439 spGlyThrLeuCysGlnProGlyAlaProAspIleCysValAlaGlyArg 455
455 ..... 455
1910 ACGGACACCTCGCGTCCAGACAGCGGTGGACATTTGCGTCACTGGCGAA 1959
456 CysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgProAs 472
472 ..... 472
1960 TGAAGCACGTGGCTCGACCGAGTCTCGGCTCGGACCTCGCGGGAGGA 2009
472 pGlyCysGlyValCysGlyGlyAspSerThrCysArgLeuValSerG 489
2010 CAAGTCGGAGTGTGTGGGTGACGCGTGTGCTGGCGAGCATCGAGG 2059
489 lyAsnLeuThrAspArgGlyGlyProLeuGlyThrGlnLysIleLeuTrp 505
2060 GCGTCTTCAGCCACGCTCACCTCGGCGCGGTACGAGGATGCTGCTGG 2109
506 IleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSer 522
2110 ATTCCTCAAGGCTCCGTCACATCTTCATCCAGGATCTGAACCTCTCT 2159
522 rAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleLeuasnGlyA 539
2160 CAGTCACTTGGCTTGAAGGACACGAGGATCCCTGCTGCTGGAGGGGC 2209
539 snTrpAlaValAspProProGlySerTyrArgAlaGlyGlyThrValphe 555
2210 TGCCTGGACCCCGCCAGCCCTGCTGCTGCTAGTGGACCACTTT 2259
556 ArgTyrAsnArgProProArgGluGluGlyGlyGlySerLeuSerAl 572
2260 CAAGTGGACAGAGGGCGCAGACAGGTC.....CAGAGCCTCGAAGC 2300
572 aGluGlyProThrThrGlnProValAspValTyrMetIlePheGlnGluG 589
2301 CTTGGGACCGGATTAATGATCTCTCATGCTGCTGCTGCTGCGCGGACCG 2350
589 luAsnProGlyValPheTyrGlnTyrValIleSerSer.ProProProIle 605
2351 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2394
605 eLeuGluAsnProThrProGluPro.....ProValProGlnL 618
2395 GACTCGGTGGCCCC...CCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2441
618 euGlnProGlu..... 621
2442 CTCGGCCCACTGTGCAGGCGGTGACCGGTCAGGCGGTGAGTGCCGCA 2491
622 .....IleLeuArgValGluProProLeuAlaProAlaProArgP 635
2492 ACCAGCTGGACAGCTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2541
635 roAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln.Me 651

```



```

360 ....ProileProArg.....CysSerGlyGluSerGluG 370
    ||| |||:
939 TAACCCCAAGCCTCAGTATGGTGGCTTATCTCTCCAGGTTCTAGCGTA 988
    ||| |||:
370 lnLeuArgAlaCysSerGlnAlaProCysProProGluGlnProAspPro 386
    ||| |||:
989 TTTATCAGCTGTCGAATTAACCCCTTGCAATGAATAATAGCTTGGATT 1038
    ||| |||:
387 ArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGlnPheMetGlyGlnLe 403
    ||| |||:
1039 CGGGCTCAACAGTGTGCAGAAATTAACAGCAAAACCTTCCGTGGATGGT 1088
    ||| |||:
403 uTyrGlnTrpGluProPheThrGluValGlnGlnGlySerGlnArgCysGluL 420
    ||| |||:
1089 CTACCAAGTGAACCCCTATACAAAGGGAAGAGGAGATCATGATGCAAC 1138
    ||| |||:
420 euAsnCysArgProArgGlyPheArgPheTyrValArgHisThrGluLys 436
    ||| |||:
1139 TGTACTGCAAGCTGAGAACTTTGAATTTTGTGCAATGTCCGGCAAA 1188
    ||| |||:
437 ValGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysValAl 453
    ||| |||:
1189 GTAAAGATGAACTCCCTCTCCCAACAAAATAATGATTTGTATGA 1238
    ||| |||:
453 aGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgA 470
    ||| |||:
1239 CGGGTTTGTGACTAGTGGATGTGATCATGATGATAGCTCTAAAGCAG 1288
    ||| |||:
470 rgProAspGlyCysGlyValCysGlyCysAspSerThrCysArgLeu 486
    ||| |||:
1289 TTTGAGATGCTTGTGGGTTTGCAGAGGTGATAATTCACTTGCAGATT 1338
    ||| |||:
487 ValSerGlyAsnLeuThrAspGlyGlyProLeuGlyTyrGlnLysII 503
    ||| |||:
1339 TATAAGGCTGTACCTCAACAGCATTAAGCAATGAATATTATCCGGT 1388
    ||| |||:
503 eLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgP 520
    ||| |||:
1389 GGTCTCTATTCAGCTGCGCCGCAAGCATCGAAATCCAGGAGCTGCAGG 1438
    ||| |||:
520 roSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleIle 536
    ||| |||:
1439 TTTCTCTCAGTTACTCGCAGTTGCAAGCTCAGTCAAAAGTATTACCTC 1488
    ||| |||:
537 AsnGlyAsnTrpAlaValAspProProGlySerTyrArgAlaGlyGlyTh 553
    ||| |||:
1489 ACCGGGGCTGGAGCATCGACTGGCTGGGGAGTTCOCCTTCGCTGGGAC 1538
    ||| |||:
553 rValPheArgTyr.....AsnArgProProArgGluGluGlyL 566
    ||| |||:
1539 CACGTTTGAATACAGCGCTCTTTCAACCGCCG..... 1572
    ||| |||:
566 ysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProValAspVal 582
    ||| |||:
1573 ....GAACGCTGTACGCGCAGGCCCCCAAAATGAGACGCTGGTCTTT 1617
    ||| |||:
583 TyrMetIlePheGlnGluGluAspProGlyValPheTyrGlnTyrValII 599
    ||| |||:
1618 GAAATTCGTATGCAAGCAAAATCCAGGATAGCTTGGAGATGATGCAT 1667
    ||| |||:
599 eSerSerProProIleLeuGluAsnProThrProGluProProValP 616
    ||| |||:
1668 T..... 1668
    ||| |||:
616 roGlnLeuGlnProGluIleLeuArgValGluProProLeuAlaProAla 632
    ||| |||:
1669 .....CCCAAGGTCATGATGAACCTCCAGCCAGCCACAAAAGA 1707
    ||| |||:
633 ProArgProAlaArgThrProGlyThrLeuGlnArgGlnValArgIlePr 649
    ||| |||:
1708 CCT..... 1710
    ||| |||:
649 oGlnMetProAlaProProHisProArgThrProLeuGlySerProAla 666
    ||| |||:

```

```

1711 .....G 1711
    ||| |||:
666 laTyrTrpIlyArgValGlyHisSerAlaCysSerAlaSerCysGlyLys 682
    ||| |||:
1712 CCTATACCTGGAGTATCGTCAGTCAGAGTGTCTCGTCTCTGTGTGGGA 1761
    ||| |||:
683 GlyValTrpArgProIlePheLeuCysIleSerArgGluSerGlyGluL 699
    ||| |||:
1762 GGTTCATATAATGTAAAGGCCATTGCTTG...CGAGATCAAAATACTCA 1808
    ||| |||:
699 uLeuAspGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProG 716
    ||| |||:
1809 AGTCATATCTCTCATCTGTCAGTGCAAAACCAAGCAGTAACTGAGCCCA 1858
    ||| |||:
716 luProCysHisGlyThrProCysProProTyrTrpGluAlaGlyGluTrp 732
    ||| |||:
1859 AAATGTGCAACGCTTCTCTCCCGGGTTACTGATGCCAGGTGAATGG 1908
    ||| |||:
733 ThrSerCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuG 749
    ||| |||:
1909 AGTACATGAGCAAGGCTGTCTGGAGGCCAGCAGCCGGAAGATCCA 1958
    ||| |||:
749 nCysArgGlnGluPheGlyGlyGlySerSerValProProGluArgC 766
    ||| |||:
1959 GTGTGTGCAAAAGAGCCCTTCCAAAGGAGGAGCAGTGTTCATTTCTC 2008
    ||| |||:
766 ysGlyHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeu 782
    ||| |||:
2009 TCTGTCCAGTGAGCACACCATCTAGGTCACAGCTGCAACAGCATGCC 2058
    ||| |||:
783 Cys...GlyHisTrpGluValGlySerProTrpSerGlnCysSerValar 798
    ||| |||:
2059 TGCCCTCCCAATGGAGCCTTGA...CCTGTCTCAGTGTTCACAGAC 2105
    ||| |||:
798 gCysGlyArgGly..... 802
    ||| |||:
2106 CTGTGACGAGGGGTGAGGAGCGTGAACCTCTCTGCAAGGGCTCTGGCG 2155
    ||| |||:
803 .....GlnArgSerArgGlnValArgCysValGlyAsnAsn 814
    ||| |||:
2156 CAGAAACCCCTCCCGAGAGAGGCGTGAACCTCTCTGCAAGGGCTCTGCC 2205
    ||| |||:
815 GlyAspGluValSerGluGlnGluCysAlaSerGlyProProGlnProPr 831
    ||| |||:
2206 GCAGAAACCCCTCCCGAGAGCAGTGTACCACT...CTCCCGAGACCTGA 2252
    ||| |||:
831 oSerArgGluAlaCysAspMetGlyProCys.....ThrThra 844
    ||| |||:
2253 GCTGACGAGGGGTGTGTCTTGGACGATGCCCAAGAACAGCCGGCTAC 2302
    ||| |||:
844 laTrpPheHisSerAspTrpSer 851
    ||| |||:
2303 AGTGGTCTGCTTCTTCTGAGGAGC 2325
    ||| |||:

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC75588

seq\_documentation\_block:

ID AAC75588 standard; cDNA; 6505 BP.

XX

AC AAC75588;

XX

DT 08-FEB-2001 (first entry)

XX

DE Human ORFX ORF1143 polynucleotide sequence SEQ ID NO:2285.

XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 cholesterol ester storage; systemic lupus erythematosus; infection;  
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 thrombosis; contraceptive; ss.

XX Homo sapiens.

XX WO2000058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PSDB; AAB41379.

XX Novel nucleic acids and peptides derived from open reading frame X,

XX useful for treating e.g. cancers, proliferative disorders,

XX neurodegenerative disorders and cardiovascular disease -

XX Claim 5; Page 1674-1678; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human OREFX open reading frames 1 to 3161. The OREFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
 CC osteoprotic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathologic conditions associated with an OREFX-associated disorder. The  
 CC nucleic acids can be used to express OREFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 6505 BP; 1154 A; 2202 C; 2005 G; 1141 T; 3 other;

alignment\_scores:

Quality:	757.50	Length:	963
Ratio:	1.698	Gaps:	42
Percent Similarity:	46.314	Percent Identity:	27.207

alignment\_block:

US-10-041-770-2 x AAC75588 ..

Align seg 1/1 to: AAC75588 from: 1 to: 6505

47 TrpGlyProTrp..... 50

||| |||||

740 TGGAGACCCCTGGTAGTAGCTGATGCCAAATGGTGAGTACCACGGACAG 789

51 .....ValGlnTrpAlaSerC 56  
 790 CCGCAGGTGAGAGCTATGTGCTGACCATCATGAACATGGTGGCTGCCT 839  
 56 ySerGlnProCysGlyValGlyVal..... 64  
 840 GTTCATGACCCCGACATGGGAACCCCATCCACATCACCATTGTGGCC 889  
 65 .....GlnArgArgSerArgThrCysGlnLeuProThrValG1 77  
 890 TGGTCCTGCTGGAAGATGAGGAGGAGGACCTAAAGATCAGCAGCATGCA 939  
 77 nLeuHisPro..... 80  
 940 GACAACACCCTGAAGAGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAA 989  
 81 ..SerLeuProLeuProProArgProProArgHisProGluAlaLeuLeu 96  
 990 GGGGATGCCCATCCCTGCACCATG..... 1015  
 97 ProArgGlyGlnGlyProArgProGlnThrSerProGlnThrLeuProLe 113  
 1016 .....ACACTGCCATCCTCTCACCAGAA..... 1039  
 113 uTyArgThrGlnSerArgGlyArgGlyProLeuArgGlyProAlas 130  
 1040 ....AGGACCTGTGTGCAGCATGACCGGCTGTGAGACCTGGGACT 1085  
 130 erHisLeuGlyArgGluGluThrGlnGluLeuArgAlaAlaArgSer 146  
 1086 GTCCCATGTGGGGCATGTGCCAGCCGCCAGCTGCAGCATCAACG 1135  
 147 ArgLeuArgAspProIleLysProGlyMetPheGlyTyrGlyArgValPr 163  
 1136 AGGACA..... 1141  
 163 oPheAlaLeuProLeuHisArgAsnArgArgHisProArgSerProProA 180  
 1142 .....CGGGCTGGCTGCTTCTCATCTGCTCCACAGCTCTGTAC 1172  
 180 rGserGluLeuSerLeuLeuSerSerArgGlyGluGluAlaIleProSer 196  
 1173 CGAGCTCGGCACAGCTTTGGCATTCACATGACGGAAGCGCAATGACT 1222  
 197 ProThrProArgAlaGluProPheSerAla.....AsnGlySerProGl 211  
 1223 GTGAGCCGTGGGAACGACCTTTTCATCTGCTCCACAGCTCTGTAC 1272  
 211 nThrGluLeuPro....ProThrGluLeuSerValHisThrProSerProG 227  
 1273 GAGCGCTGCCCTCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1322  
 227 InAlaGluProLeuSerProGluThrAlaGlnThrGluValAlaProArg 243  
 1323 GTTCCTGACCGTGGTGGGCTGTGCTGACGACGACCTCTGCTGCAAGG 1372  
 244 ThrArgProAlaProLeuArgHis.....Pr 253  
 1373 ACATTATCGACTTCCCTCGGTGCGCACCTGCGCTCTCTATGATGTAAGC 1422  
 253 oArgAlaGlnAlaSerGlyThrGluProProSerPro..... 265  
 1423 CACAGTGGCGCTCCAGTACGGGGCTCTACTGCTTCTGCGAGGACAT 1472  
 266 .....ThrHisSerLeuGlyGlyGlyPhePheArgAla 277  
 1473 GGATAATGTCTGCCACACACTCTGTGTGTGGGA..... 1510  
 278 SerProGlnProArgProSer...SerGlnGlyTrpAlaSerProGl 293  
 1511 .....CACCTGTCTCCACAGCTGGATGCTGTGGAGCGGACCGCG 1554  
 293 nValAlaGlyArg.....ArgProAspPropheP 303

```

1555 TGTG...GGGAGATAAAGTGTGTCTAGTGGGAGTCCGTACCCGTGGG 1601
|||||
303 roSerValProArgGlyArgGlyGlnGlnGlyProTrp..... 317
|||||
1602 CTTCCGGCCGAGG.....CCGTGGATGTG 1627
318 ..GlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProG1 333
|||||
1628 GCTGTGTGTGGTGGAGCCCT...GGTCCA..... 1654
333 nHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSerS 350
|||
1655 .....TCTGCTCAGGAGGTG 1670
350 exLeuTrpSerLeuPheAlaProSer...Ser.....ProIleProA 363
|||||
1671 TGGCATGGGCTACAGAGCGCGAGCGCAGTGCAGCGAGCTACGCCCA 1720
363 rg.....CysSerGlyGlySerGluGlnLeuArgAlaCys 374
|||
1721 AATACAAAGGCAGATACTGTGTGGGTGAGCGCAAGCGCTTCGGCTCTGC 1770
375 SerGlnAlaProCysProProGluGlnProAspProArgAlaLeuGlnCy 391
|||
1771 AACCTGCAGGCGCTGCTGCTGCTGGCGCCCTCTCCCGCAGTCCAGTG 1820
391 sAlaAlaPheAsnSerGlnGlnPheMetGlyGlnLeuTrpGlnP 408
|||
1821 CAGCCACTTTCAGCTATGCTCTACAAAGGCGCAGTGCACACATGGGTG 1870
408 roPheThrGluValGlnGlnSerGlnArgCysGluLeuAsnCysArgPro 424
|||
1871 CGGTG.....GTCAATGAGCTGAACCCCTGGAGCTGCACCTGCCCGGCC 1914
425 ArgGlyPheArgPheTrpValArgHisThrGluLysValGlnAspGlyTh 441
|||
1915 GCGATGAGTACTTTCCAAAGAGTTCGGGACGCGCTGGTGCATGGCAC 1964
441 rLeuCysGlnPro.....GlyAlaProAspIleCysValAlaGlyArgC 456
|||
1965 CCCTGCTACCAAGGTCGAGCCGAGCGGAGCCTCTGCATCAACGGCATCT 2014
456 ysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgProAsp 472
|||
2015 GTAAGACGTGGCTGTGACTTCGAGATTGACTCCGGTCTATGAGGNC 2064
473 GlyCysGlyValCysGlyGlyAspSerThrCysArgLeuValSerGl 489
|||
2065 CGCTGTGTGTGTGCCACGCAACGCTCCACTGCCACACCGTGGCGG 2114
489 yAsnLeuThrAspArgGlyGlyProLeuGlyTrpGlnLysIleLeuTrpI 506
|||
2115 GACCTTCNAGAGGAGCCGAGGTCWNGGGTATGTGGATGTGGGCTGA 2164
506 leProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgProSerSer 522
|||
2165 TCCACGCGGCGCAGCGAGATCCGATCCAAAGAGGTTCGCGAGGTGCC 2214
523 AsnTrpLeuAlaLeuArgGlyProGly...GlyArgSerIleIleAsnGl 538
|||
2215 AACTCTCGGCACTGGGAGCGAGGACCCGAGAGACTTCTCTCAATGG 2264
538 yAsnTrpAlaValAspProGlySerTrpArgAlaGlyGlyThrValP 555
|||
2265 TGGCTGGACCATCCAGTGAACGGGACTACCAAGGTGGCAGGACCACT 2314
555 heArgTrpAsnArgProProArgGluGluGlyLysGlyGluSerLeuSer 571
|||
2315 TCACATACGACGC.....AGGGCAACTGGGAGAACCTCAGC 2352
572 AlaGluGlyProThrThrGlnProValAspValTrpMetIlePheGlnGl 588
|||

```

```

2353 TCCCGGGTCCCAAGAGGAGCGCTGTCTGGATC...CAGGTGCTGCCTC 2399
588 uGluAsnProGlyValPheTrpGlnTrpValIleSerSerProProI 605
|||||
2400 CCGTGGCCCAAGCGGGGGAGAGCGGAGTCCCAAGCCAGCACCC 2449
605 leLeuGluAsnProThrProGluProProValProGlnLeuGlnProGlu 621
|||
2450 TCCATGSCAGGTCTCGTCTGGAGGAGTGCCTGTGTTCAATC... 2493
622 IleLeuArgValGluProProLeuAlaProAlaProArgProAlaArgTh 638
|||||
2494 .....ACAGAGCTGGCTGTGAGCCAGGCCCT...CCTGCTCGGC 2531
638 rProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProP 655
|||||
2532 CTCTACCTCAGTTCCCATCTTTAAATGGCCCAATCTTGTAGCTGCAG 2581
555 roHis..... 656
2582 TTCACAGAGGTGGTGGGTCAAGCTCCTTTAGGACTGGGTGGATGGAGA 2631
657 .....ProArgThrProLeu..... 661
2632 AGACACCTGTGCTCATGGCCCGCCGCTGCCACCCAGCTGTGTCCA 2681
662 ....GlySerProAlaAlaTrpTrpLysArgVal..... 671
2682 GGAGAGCAACCTGGGTGCTACAGTACACCATCCACAGGAGSCAG 2731
672 ..GlyHis..... 673
2732 GTGGCCAGGAGAGGTCCGCGCCCGCTGTCTCTGGCATTATGGGCC 2781
674 ...SerAlaCysSerAlaSerCysGlyValTrpArgProIlePh 689
|||
2782 TGGACCAAGTCCACAGTCCACCTGGCGGAGAGTGTGCAGAGCGACAATGT 2831
689 eLeuCysIleSerArgGluSerGlyGluLeuAspGluArgSerCysA 706
|||
2832 GTACTGTGGAGCGGAGGAGG...CCGTGGACGAGGACACTGTG 2878
706 laAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrPro 722
|||
2879 ACCCCTGGCGCGCTGTATGACCAACAGAGGAAGTGCAGCAGCAGCC 2928
723 CysProTrpTrpGluAlaGlyGluTrpThrSerCysSerArgSerCy 739
|||
2929 TCCCTGCGCAGGTGGTGGCAGGTGAGTGGCAGCTGTCTCCAGCTCTG 2978
739 sGlyPro...GlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPhe 755
|||
2979 CGGCGCTGGGCGCTCTCCCGCGGCGCTGTCTGTCATCCGAGCGTGG 3028
755 ly...GlyGlyGlySerSerValProProGluArgCysGlyHisLeuPro 770
|||
3029 GGCTGGATGACGAGAGCGCTCCAGCCCGCTGTGACACCTTCCC 3078
771 ArgProAsnIleThrGlnSerCysGlnLeuArgLeu...Cys...GlyH 785
|||
3079 CGGCCCCCTACTGAACCCCTTGCAACCGCCATGTACCTGTCCCGGCCAC 3128
785 strGluValGlySerProTrpSerGlnCysSerValArgCysGlyArg 802
|||
3129 CTGGCTGTGGGGAAC...TGGTCTCAGTGTGACATGTGGGAGG 3175
802 lyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluVal 818
|||
3176 GCACACGCGCGGAATGTCTCTGCACCAATGACCGGTGTCCCTGT 3225
819 SerGluGlnGluCysAlaSerGlyProProGlnProProSerArgGluAl 835
|||
3226 GACGAGGCCAG.....CAGCCAGCGCAGGAAGTCAC 3257

```

```
835 acysaspMetGlyProCysThr.ThrAlaTrpPheHisSerAspTrpSer 851
:||||:||||: ||| ||| |||:|
3258 CTGCTCTCGCACTCTGTGGTGGCCCTGGGCACAC.....TGGGCC 3301

852 SerLysVal...SerProGluProProAlaIleSer 862
||||:|:| ||||| |||
3302 CTGAAGGCTCAGGACGAGGCTCTCCAGCCACGAGC 3337
```





```

|||||
578 GTCTCGGGAAGGGTGGCCACTTCGAGCTCCCGCTTCCACCTAGGGA 627
|||||
134 rgGluGluThrGlnGluLeuAlaAlaArgArgSerArgLeuArgasp 150
|||||
628 GAGAGGAGACCCAGGAGATTCAGAGGGCCAGGAGGTCCGGCTTCGAGAC 677
|||||
151 ProIleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuPr 167
|||||
678 CCATCAGCCAGGATGTTGGTATATGGGAGGTGCCCTTGTGATTGCC 727
|||||
167 OLeuHisArgAsnArgArgHisProArgSerProProArgSerGluLeuS 184
|||||
728 ACTGCACCGGAACCGCAGCACCCCTCGGAGGCCACCCAGATCTGAGCTGT 777
|||||
184 erLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProArg 200
|||||
778 CCTGATCTCTTCTAGAGGGGAAGAGCTATTCCGTCCCTACTCCCAAGA 827
|||||
201 AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh 217
|||||
828 GCAGAGCCATTCTCGCAACCGCAGCCGCCCAAACTGAGCTCCCTCCAC 877
|||||
217 rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG 234
|||||
878 AGAATCTGTCTCCACACCCCATCCCTCCCAACAGCAACTCTTAAGCCCTG 927
|||||
234 luThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg 250
|||||
928 AAATCTCTCAGACAGAGGTGGCCCCCAGAACCCAGGCTGCCCCCTACGG 977
|||||
251 HisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHi 267
|||||
978 CATCACCCAGAGCCAGGCCCTCTGGCACAGAGCCGCCCTCACCCACGCA 1027
|||||
267 sSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArgArgP 284
|||||
1028 CTCCTTAGGAGAAGTGGCTTCTTCGGTGCATCCCTCAGCCACCAAGGC 1077
|||||
284 roSerSerGlnGlyThrAlaSerProGlnValAlaGlyArgArgProasp 300
|||||
1078 CAAGTTCACAGGTTGGGCCAGTCCCCAGGTAGCAGGGAGAGCCGCTGAT 1127
|||||
301 ProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProfr 317
|||||
1128 CTTTTCCTTCGGTCCCTCGGGCCGAGGCCAGCAGGSCCAAGSGCCTTG 1177
|||||
317 pGlyThrGlyThrProHisGlyProArgLeuGluProAspProGlnH 334
|||||
1178 GGGACCGGGGGGACTCTCTCAGGGCCCGCTGGAGCCTGACCCCTCAGC 1227
|||||
334 isProGlyAlaThrLeuProLeuLeuSerAsnGlyProHisAlaSerSer 350
|||||
1228 ACCCGGCGCTGGCTGGCCCTGCTGAGCAACGGCCCATCCAGCTCC 1277
|||||
351 LeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyGl 367
|||||
1278 CTCTGAGGCTCTTCTGCCAGTAGCCCTATTCCAAAGATGTTCTGGGGA 1327
|||||
367 uSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGluGlnP 384
|||||
1328 GAGTGAACAGCTAAGAGCTGGAGCCGAGCCAGCGCCCTGCCCTTGAGCAGC 1377
|||||
384 roAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMet 400
|||||
1378 CAGACCCCGGGGCCCTGAGTGGCAGCCTTTAACTCCCGAGGAATTCAATG 1427
|||||
401 GlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGlnGlySerGlnAr 417
|||||
1428 GGCAGCTGTATCAGTGGAGGCCCTTCACTGAAGTCCAGGCTCCAGCG 1477
|||||
417 gCysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArgHisT 434
|||||

```

```

1478 CTGTGAACCTGAACCTGCGCGCCCGCTTCCGCTTCTATGTATCGGTCA 1527
|||||
434 hrGluLysValGlnAsnAspGlyThrLeuCysGlnProGlyAlaProAspIle 450
|||||
1528 CTGAAAAGGTCCAGGATGGAGCCCTGTGTGAGCTGGAGCCCTGACATC 1577
|||||
451 CysValAlaGlyArgCysLeuSerProGlyCysAspGlyIleLeuGlySe 467
|||||
1578 TGTGTGCTGGAGCTGTCTGAGCCCGGCTGTGATGGGATCCTTGGCTC 1627
|||||
467 rGlyArgArgProAspGlyCysGlyValCysGlyGlyAspAspSerThrC 484
|||||
1628 TGGCAGGCTCTGTATGCTGTGGAGTCTGTGGGGTGTGATTTCTACCT 1677
|||||
484 ysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyr 500
|||||
1678 GTCCGCTTGTTCGGGGAACCTCACTGACCGAGGGGCCCTTGGGCTAT 1727
|||||
501 GlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGl 517
|||||
1728 CAGAAGATCTTGTGATTCAGCGGGAGCCTTGGGCTCCAGATTCGCCA 1777
|||||
517 nLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgS 534
|||||
1778 GCTCCGGCTAGTCCAACTACCTGGCACCTTCGTGGCCCTGGGGGCGGT 1827
|||||
534 erIleIleAsnGlyAsnTrpAlaValAspProProGlySerTyrArgAla 550
|||||
1828 CCATCATCAATGGGAACCTGGGCTGTGGATCCCTGGGTCTCTACAGGCC 1877
|||||
551 GlyGlyThrValPheArgTyrAsnArgProProArgGluGluGlyLysGl 567
|||||
1878 GCGGGGACCGTCTTTCGATATACCCGCTCCCGAGGAGGAGGCAAGG 1927
|||||
567 yGluSerLeuSerAlaGluGlyProThrThrGlnProValAspValTyrM 584
|||||
1928 GGAGAGTCTGCGCTGAAAGGCCCCACACCCAGCCCTGTGTGATGTCTATA 1977
|||||
584 erIlePheGlnGluAsnProGlyValPheTyrGlnTyrValIleSer 600
|||||
1978 TGATCTTTCAGAGGAACCCAGCGCTTTTATCAGTATGTCATCTCT 2027
|||||
601 SerProProIleLeuGluAsnProThrProGluProProValProGl 617
|||||
2028 TCACCTCTCCAATCTTGAGAACCCACCCAGAGCCCTGTGCCCA 2077
|||||
617 nLeuGlnProGluIleLeuArgValGluProProLeuAlaProAlaProA 634
|||||
2078 GCTTCAGCGGAGATTCAGGGTGGAGCCCCACTTGTCCCGGACCCC 2127
|||||
634 rgProAlaArgThrProGlyThrLeuGlnArgGlnValArgIleProGln 650
|||||
2128 GCCAGCCCGGACCCAGGCCACCTCCAGGCTCAGGTCCGGATCCCCCAG 2177
|||||
651 MetProAlaProProHisProArgThrProLeuGlySerProAlaAlaTy 667
|||||
2178 ATGCCCGCCCGCCCATCCAGGACCCCTGGGGTCTCCAGCTCGGTA 2227
|||||
667 rTriPlysArgValGlyHisSerAlaCysSerAlaSerCysGlyLysGlyV 684
|||||
2228 CTGGAACAGAGTGGACACTCTGCATCTCAGCTCCTCGGGAAAGGTG 2277
|||||
684 alTrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeu 700
|||||
2278 TCTGGCGCCCATTTTCTCTGTCATCTCCCGTGTGAGTCGGGAGAGGAATG 2327
|||||
701 AspGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluPr 717
|||||
2328 GATGACCGAGCTGTGCGCGGGTGCAGGCCCCCGCCCTCCCTGACCC 2377
|||||
717 oCysHisGlyThrProCysProProTyrTrpGluAlaGlyGluThrS 734
|||||
2378 CTGCCAGCGCACCCCATGCCCCCATACTGGGAGGCTGGCGAGTGGACAT 2427
|||||

```



80203 AGGGCCCCGAAGGTGCTCGGGACCTTGGGTCCAGTGGGCCCTCTTGCTCC 80252  
58 GlnProCysGlyValGlyValGlnArgArgSerArgThrCysGlnLeuPr 74  
80253 CAGCCCTGGGGGTGGGTGGGTGAGCGCAGGAGCGGACATGTGACGTCC 80302  
74 oHrValGlnLeuHisProSerLeuProLeuProProArgProProArgH 91  
80303 TACAGTGCAGCTCCACCCAGGTGCCCCCTCCCTCCCTCCCGGCCCCCAAGAC 80352  
91 isProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThrSer 107  
80353 ATCCAGAACCCCTCTCCCGCGGGCCAGGGTCCAGACCCAGACTTCT 80402  
108 ProGluThrLeuProLeuTyrArgThrGlnSerArgGlyArgGlyGlyPr 124  
80403 CCAGAAACCCCTCCCTTTACAGGACACAGTCTCGGGGAAGGGGTGGCCC 80452  
124 oLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluLea 141  
80453 ACTTCGAGGTCCCGCTTCCACCTAGGGAGAGAGAGACCCAGGAGATTC 80502  
141 rgAlaAlaArg..... 144  
80503 GAGCGGCAGGAGGTGAGAGCCTGGGTGGAGAGGTGGCCTTGGGCAA 80552  
144 ..... 144  
80553 GGTGGGGCTTAGCATGAAGAAAGGGAGCACCGGTGGCTTCAGAGGC 80602  
144 ..... 144  
80603 TTGTCTTTGGTGCCAGGAAGTAACTGGGAAACAGGGTCAGAGGTGGTC 80652  
145 ..... Arg\_SerArgLeuArgAs 150  
80653 AGACTGTGCCCCACATCTGAACACCTCTCAGGTCCCGGTTCGAGA 80702  
150 pProileIysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuP 167  
80703 CCCCATCAAGCCAGGAATGTTCGGTTATGGAGAGTGCCTTTGCAATTGC 80752  
167 roLeuHisArgAsnArgHisProArgSerProProArgSerGluLeu 183  
80753 CACTGCACCGGAACCGGAGCACCTCGGAGCCCCACCATCTGAGCTG 80802  
184 SerLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProAr 200  
80803 TCCCTGATCTCTAGAGGGAGAGGCTATCCGTCGCCCTACTCCAAG 80852  
200 gAlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProT 217  
80853 AGCAGAGCCATTCTCCGAAACGCGACGCCCCCAAACTGAGTCCCTCCCA 80902  
217 hrGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerPro 233  
80903 CAGAACTGTCTGTCCACCCCTCCCTCCCGGAGAGAACTCTAAGCCCT 80952  
234 GluThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuAr 250  
80953 GAAACTGTCTCAGACAGAGGTGGCCCCCAGAACACAGGCTGCCGCCCTAGC 81002  
250 gHisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267  
81003 GCATACCCCGAGCCAGGCTCTGGGCACAGAGCCCGCTCACCACAGC 81052  
267 isSerLeuGlyGlyGlyPhePheArgAlaSerProGlnProArgArg 283  
81053 ACTCCTTAGAGAGAGGTGGCTTCTCCGTGCATCCCTTCAGCCACGAAAG 81102  
284 ProSerSerGlnGlyThrAlaSerProGlnValAlaGlyArgArgProAs 300  
81103 CCAAGTTCACAGGTGGCGTGGAGATCTGTCCAGCCGCTGACAGCCAGGTGG 81152

300 pProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProT 317  
81153 TCCTTTTCTTCGTTCCTCGGGGCCAGGCGCAGCAGGGCCAAAGGCCCTT 81202  
317 rpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProGln 333  
81203 GGGNACGGGGGGGACCTCTCACGGGCCCGCCCTGGAGCTGACCCCTCAG 81252  
334 HisProGlyAlaIleProLeuProLeuLeuSerAsnGlyProHisAlaSerSe 350  
81253 CACCGGGCGCTGCTGCCCTGTGAGCAACGGCCCCCATGCCAGCTC 81302  
350 rLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyG 367  
81303 CCTCTGGAGCCTCTTGTCTCCAGTAGCCCTATTCCAAGATGTTCTGGGG 81352  
367 luserGluGluLeuArgAlaCysSerGln..Ala..... 377  
81353 AGAGTGAACAGCTAAGAGCCTGACGCCAAGCGGTGAGTCTCTCTCGGGCCT 81402  
377 ..... 377  
81403 CCCCTCCCAACCCCGACCTCCAGTGTGCTTCCCTGCCCTGAAGCTGGGT 81452  
377 ..... 377  
81453 CTTACAGTTCCGCTTGGCACCTGAGGAGAGAGGCCCTTGGCATCTGACCA 81502  
377 ..... 377  
81503 COTCAGGCGAGGGTCTTGGAGCTTTCGGTGACCTTCCGCTGAGCTGAGCCTCCACCT 81552  
377 ..... 377  
81553 GCTCCCGAGTTTCAGCCCTGCCCCCTACCTCATTTTGTCTCCCGACGCTG 81602  
378 ..... ProCysProProGluGlnProAspPr 386  
81603 ACTCCTTTGTACCCCTCACCAGCCCTGCCCCCTGAGCAGCCAGACCC 81652  
386 oArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnL 403  
81653 CCGGGCCCTGCAGTGGCAGCCTTAACTCCAGGAATTCATGGGCCAGC 81702  
403 eutyrglnTrpGluProPheThrGlu.Val..... 412  
81703 TGTATCAGTGGAGCCCTTCACTGAAGGTGAGGTTCCTGTGCTCACCCCTG 81752  
412 ..... 412  
81753 GGCAGTGGTGGCTTGAATTGGGATGAAGGGAGAGAGATACCTGCTT 81802  
412 ..... 412  
81803 ACTCCAGCCCTGAATGACTTCCAGCCCTCTGCTTCCCTCGGCCATGC 81852  
412 ..... 412  
81853 CTTCCTTCTCCTCCCTGGGGCTGGGTGAGGACACACAGGTGGTGTAGT 81902  
412 ..... 412  
81903 GGTCTGCAGAAAGGTGCGGGTGGGAGGAGGTGTGGAGACACGCTTT 81952  
412 ..... 412  
81953 GTCCGAGTCCCTGGGAAGGCCATCCTCTGGGGGTGAGTGGCTGTGACA 82002  
412 ..... 412  
82003 CAAGAGGGCCATGTTGGAGAGATCTGTCTCAGCCGCTGACAGCCAGGTGG 82052

412 ..... 412  
82053 GGGTGTGCCAGCATCTGGGCACGTGCTGATATCGGTGGTGTCTCCAGGTTAC 82102  
412 ..... 412  
82103 CATCCGCTTCATTTTAGCGCTGTGTTAACTTGACACCGGGATAAGCACATC 82152  
412 ..... 412  
82153 AAGAGCTCTGAACACACACAGAATCATGGGCATCAGGGAATGCTCTCTCTC 82202  
412 ..... 412  
82203 AGTTCAGATCCCTTGATGGTCTCGGTGACCTCACTGACCTGACCACC 82252  
412 ..... 412  
82253 TTTTTTTTTTTTCTGAGACAGAGTCTCACTCTGTCTACCCAGGCTGGA 82302  
412 ..... 412  
82303 GTACAGTGGCAGATCTCAGCTCACTGCAACCTCTGCTTCCGGGTTCAA 82352  
412 ..... 412  
82353 GCAATCTCTGCTGCTCAGCTCCCGAGTAGCTGGGACTACAGGCATGTC 82402  
412 ..... 412  
82403 CACCACCCCTGGCTAATTTTGTATTGTTAGTGGAGATGGGTTTCAACA 82452  
412 ..... 412  
82453 CGTTGGCCAGGCTGCTCTGAACCTCTGACCGAATGATGCACCCACCTCG 82502  
412 ..... 412  
82503 GCCTCCAGAGTGTGGGATTACAGGCATGAGCCACTGCGCCCGGCCCTG 82552  
412 ..... 412  
82553 ACCACCTCAGCTTGTGCTGTGTGCATATTTCCCTTCATCCACAGTGACCTG 82602  
412 ..... 412  
82603 GGCAACCTCAAGTGCCTCTGGGGCCCTCTCTCAGCTAACCTCCACACC 82652  
413 ..... GlnGlySerG 416  
82653 AGGAGGCCCACTACCACCTCTTACCTCTGCTTGTCCCTTAGTCCAGGGCTCCC 82702  
416 lnArgCysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArg 432  
82703 AGCGCTGTGAACCTGAACCTGCGCGGCCGCTGCTCCGCTTCTATGTCCGT 82752  
433 HisThrGluValGlnAspGlyThrLeuCysGlnProGlyAlaProAs 449  
82753 CACACTGAAAGAGTCCAGGATGGGACCTGTGTGAGCTGGAGCCCTGA 82802  
449 pIleCysValAlaGlyArgCysLeu..Ser. 458  
82803 CATCTGTGTGCTGGACGCTGCTGTGGTGGGGAAGACAGAGTGTGTGTGC 82852  
458 ..... 458  
82853 ACACACATGCATATGCACACACACATGCCCCCATATGCATACACAT 82902  
458 ..... 458  
82903 GTACACATATGTATGACACATGCACACATGCAAGCACATACACACAC 82952  
458 ..... 458  
82953 GCATATGGCACAGACACATGCACACACGACACACACACACACACACACAC 83002  
458 ..... 458  
83003 ACGCATATGCACACACATGCACATGCATATGCACACACATGCATGCACAT 83052  
458 ..... 458  
83053 GTGCACACGTGCATGAACACATGCACACACACGTATGCAGACACATGCAC 83102  
458 ..... 458  
83103 ACACATGTAGACACACATGCATGAACACATGCACACATGCACACACACAGT 83152  
458 ..... 458  
83153 ATTTGCACAGTCTCAATTCAACACAAACACCATACTGAGATCACTGAG 83202  
458 ..... 458  
83203 GGAGAATTCTCACCTGCCAAGCTGCGCTGAGCAGAGCCTGAAGGATCTGA 83252  
458 ..... 458  
83253 TCGGCACCTGTCCATGTCCCTGGCTCTGGTGGGATGGTGGGCTGTT 83302  
458 ..... 458  
83303 TTTGTGCTCTCACTTGTGSCACAAAAGCAGGTAGTGAGCTGAGGCTCC 83352  
458 ..... 458  
83353 CGAGGGACCGGGTGGGTTGAGTGTGCTGCTGGCTTCTGTGGCCACT 83402  
459 ..... ProGlyCysAspGlyIleLeuGlyse 467  
83403 GCCTCACCTCACTCTCCAGAGCCCGGCTGTGATGGGATCCTTGGCTC 83452  
467 rGlyArgArgProAspGlyCysGlyValCysGlyGlyAspSerThrC 484  
83453 TGGCAGGCTCCTGATGGCTGTGGAGTCTGTGGGGTGATGATTTACT 83502  
484 ysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyr 500  
83503 GTGCGCTTCTTTCGGGAACCTCACTGACCGAGGGGCCCTGGCTAT 83552  
501 GlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAla 517  
83553 CAGAAGATCTTGTGATTCCAGCGGAGCCTTGGGCTCCAGATTGCCCA 83602  
517 nLeuArgProSerSerAsnTyrLeu. 525  
83603 GCTCCGCTTAGCTCCAACTACCTGGGTGAGCACCAGCTGCCCTCCCTT 83652  
525 ..... 525  
83653 CCACITTCGCTCTGTTCGCGCCCTCCATACCCCTACTCAGACAGTGCAGC 83702  
525 ..... 525  
83703 AAGCCAAACAGGGGGAAGTCCAGGGCCTAGCCCTCCCTCTGGGAAGCA 83752  
525 ..... 525  
83753 GTGAGGAAGCTGAGAGGCTTGGGGGATCTTAGTTCGTGGGAGCTT 83802  
525 ..... 525  
83803 CTATAGGCTAAGCACACGGTGTGGGAGGAGGATATTATCACCCCTGGA 83852  
526 ..... Ala..LeuArgGlyProGlyGly 532  
|||||

83953	ATTTC	CGA	TCTCT	CA	CTCTG	ACCCG	CAC	GA	CACTTC	GTGG	CCCTGG	GGGC	83902			
533	ArgSer	Ile	Ile	Asn	Gly	Asn	Trp	Ala	Val	Asp	Pro	Gly	Ser	Tyr	Arg	549
83903	CGGT	CCAT	CA	TCA	TGG	GA	ACTGG	GTG	TG	ATC	CCCTG	GGT	CCT	TAC	AG	83952
549	GAla	Gly	Val	Thr	Val	Phe	Arg	Tyr	Asn	Arg	Pro	Arg	Glu	Glu	Gly	566
83953	GGCC	GGG	GAC	CGT	CTT	CGA	TATA	ACC	GTCT	CTCC	CAG	GAG	GGCA	84002		
566	YsGly	Glu	User	Leu	Ser	Ala	Glu	Gly	Pro	Thr	Gln	Pro	Val	Asp	Val	582
84003	AAGG	GAG	AGT	GTG	TCG	GTGA	AGCC	CC	ACC	ACC	CA	CGT	GTG	GTG	ATC	84052
583	Tyr	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	583		
84053	TATG	TG	AGC	CTT	GGG	CC	AGG	GC	AGC	TGA	ATG	CTGG	GAG	GGC	TGT	84102
583	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	583		
84103	CCC	TCT	GGG	C	AG	ACT	GTG	GTG	TG	CA	AG	ATG	GGG	AG	AG	84152
584	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	Me	584	
84153	TCC	TCT	TCT	GG	CC	AC	CC	CC	CA	TAT	TAT	CTT	CT	TCT	CT	84202
584	Ile	Phe	Gln	Glu	Asn	Pro	Gly	Val	Phe	Tyr	Gln	Tyr	Val	Ile	Ser	601
84203	GAT	CTT	C	AG	G	A	A	A	ACC	C	AG	G	G	T	T	84252
601	er	Pro	Pro	Pro	Ile	Leu	Glu	Asn	Pro	Thr	Pro	Glu	Pro	Pro	Val	617
84253	CAC	CT	CT	CT	CA	TCT	CT	CA	TCT	CT	CT	CT	CT	CT	CT	84302
618	Leu	Gln	Pro	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	620		
84303	CTT	C	AG	CC	G	CG	GT	A	AG	CT	CT	C	AG	CC	CT	84352
620	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	620		
84353	GCT	G	C	AG	GG	CT	GG	CT	GG	CT	GG	CT	GG	CT	GG	84402
621	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	634		
84403	CCT	G	C	A	C	A	G	A	T	CT	G	A	G	G	T	84452
635	Pro	Ala	Arg	Thr	Pro	Gly	Thr	Leu	Gln	Arg	Gln	Val	Arg	Ile	Pro	651
84453	CC	AG	CC	CC	GG	AC	CC	CT	CC	AG	GT	C	AG	T	CC	84502
651	t	Pro	Ala	Pro	Pro	His	Pro	Arg	Thr	Pro	Leu	Gly	Ser	Pro	Ala	668
84503	GCC	CC	CC	CC	GG	CC	AT	CC	C	AG	G	AC	CC	CT	GG	84552
668	rp	Lys	Arg	Val	Gly	His	Ser	Ala	Cys	Ser	Ala	Ser	Cys	Gly	Val	683
84553	GG	AA	AC	G	AG	T	GG	AC	T	CT	G	AT	G	CT	CT	84602
683	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	683		
84603	AC	A	T	C	A	G	A	G	T	CC	CG	CA	T	CT	C	84652
683	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	683		
84653	C	O	C	G	C	A	T	CT	GG	G	C	C	A	C	CC	84702
683	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	683		
84703	CT	G	A	G	CT	GG	G	CT	A	G	CT	CT	CT	GG	CT	84752
684	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	698		
84753	C	A	G	T	CT	GG	C	C	C	A	T	T	T	CT	CT	84802

699	GluLeuAspGluArgSerCysAlaIaGlyAlaArgProProIaSerPr	715
715		
8403	GAACCTGGATGAACGACAGCTGTGCCGCGGGTGCACAGGCCCGCTCC	84852
715	oGluProCysHisGlyThrProCysProPro.....	725
84853	TGAACCTGTCCACGGCACCCCATGCCCCCATATGAGTAGTGGGGAGCC	84902
725		
84903	CACGGGGAGGTTAGGTACTGGAACACACAGCAGTTGCCCCAGAAATCTT	84952
725	.....	725
84953	ACCTGAACCTGCACCTCAACGCAACATCTCTCTGGCTGCAGAGACACA	85002
725	.....	725
85003	TTAGATTAGAAGCAGACATTCGGTAGGCAGAAAGCAAGCATGGCAGG	85052
725	.....	725
85053	CTGAGCCCCCAGGAACCCAGACTCCAAACGCCAAAGTGCCTCTCC	85102
725	.....	725
85103	TCTGAGCCCCCAGCTCCAGTGTGTCTTCCATCTCTGTCACAGGCAGC	85152
725	.....	725
85153	CCACACATCTCATATGTCTCGGCTTCTCCCTGCTGCTGGTGCCT	85202
726	..TyrTrpGluIaGlyIuTrpHrSerCysSerArgSerCysGlyPro	741
85203	GCAGCTGGGAGCTGGCAGTGGACATCCTCAGCCGCTCTCTGGGCC	85252
742	GlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGlyGly	758
85253	GGCACCCAGCACCGCCAGCTGCAGTCCGCGAGGAATTTGGGGGGGTG	85302
758	ySerSerValProProGluArgCysGlyHisLeuProArgProAsnIle	775
85303	CTCCCTGGTGGCCCCGGAGCGGTGTGGACATCTCCCGCGGCCCAATCA	85352
775	hrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluValGlySerPro	791
85353	CCAGTCTTGCCAGCTGGCCTCTGTGCCATTTGGAAATTTGGGTCTCCT	85402
792	Trp.....	792
85403		
792	TGGAGCCAGTTCAGTTGCCAGCAGAGTGTCTGGGAGGGGAATGG	85452
85453	GCACAGGTGAACAACAGCAGTGGTTTTTCAACTCTTTTTTTCATCAGCAGA	85502
792	.....	792
85503	AACATATTACAACAGGTTTCATATGGACCCCAATATAGATCAGAT	85552
85553	AAGGACTGAACAGGACTGGGGGCCAGGCCCTAGGGCTCAGCCTCC	85602
792	.....	792
85603	TCCTACGCCACGAAGCCATGTACTGTCTGGCTGGGATCGAAGGCAGA	85652
792	.....	792
85653	GCCTGGTCTCTGAAGCCTAGAGCTGAAGCTGGCTCAGAAAGGTCTGGGAG	85702

```

793 .....Ser..GlnCysSerValArgCysGly 800
85703 CCCAGCTCCCTGGATTCCCTCGCCCTCAGTCTCGGTGCGTGGCC 85752

801 ArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspG1 817
|||||
85753 CGGGCCAGAGAAGCGGAGGTCCTGCTGTGGGAACAAATGGTGATGA 85802
|||||
817 uValSerGluGlnGluCysAlaSerGlyProGlnProGlnProSerArgG 834
|||||
85803 AGTGAGCAGCAGGAGGTGGGTGAGGCCGCCCGCCAGCCCGCCAGCAG 85852
|||||
834 luAlaCysAspMetGlyProCysThrAlaTrpPheHisSerAspTrp 850
|||||
85853 AGGCCTGTGACATGGGGCCCTACTACTGCTGCTGCTTCCACAGCGACTGG 85902
|||||
851 SerSerLysValSerProGluProProAlaIleSerCysIleLeuGlyAs 867
|||||
85903 AGCTCCAGGTGAGCCGCGAACCAGCCCATATCTGTCATCTCTGGTAA 85952
|||||
867 nHisAlaGlnAspThrSerAlaPheProAla 877
|||||
85953 CCACGCCAGGACACCTCAGCCTTTCCAGCA 85983

seq_name: gb_htg:AC092479

seq_documentation_block:
LOCUS AC092479 142023 bp DNA linear HTG 05-FEB-2002
DEFINITION Mus musculus clone rp23-231115 strain C57BL/6J, WORKING DRAFT
SEQUENCE 3 unordered pieces.
ACCESSION AC092479
VERSION AC092479.12 GI:18497125
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Jia.H., Zhang.P., Lin.S., Wu.H. and Roe.B.A.
Mus musculus BAC Clone rp23-231115
Unpublished
Jia.H., Zhang.P., Lin.S., Wu.H. and Roe.B.A.
2 (bases 1 to 142023)
Direct Submission
Submitted (10-JUL-2001) Department of Chemistry and Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Feb 5, 2002 this sequence version replaced gi:18390246.
----- Genome Center
Center: Department of Chemistry and Biochemistry
The University of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 32949: contig of 32949 bp in length
* 33049: gap of unknown length
* 33050 78918: contig of 45869 bp in length
* 78919 79018: gap of unknown length
* 79019 142023: contig of 63005 bp in length.
* Location/Qualifiers
* 1. 142023
* /organism="Mus musculus"
* /strain="C57BL/6J"
* /db_xref="taxon:10090"
* /clone="rp23-231115"
* /clone_lib="RPCI mouse BAC library 23"
FEATURES
Source

```

```

BASE COUNT 37523 a 34071 c 31510 g 38713 t 206 others
ORIGIN

alignment_scores:
  Quality: 2670.50      Length: 1750
  Ratio: 3.826         Gaps: 28
  Percent Similarity: 39.886 Percent Identity: 35.371

alignment_block:
US-10-041-770-2 x AC092479 ..

Align seg 1/1 to: AC092479 from: 1 to: 142023

28 LeuSerGlyHisSerLeuGlnThrProThrGluGluGlnGlyProG1 44
|||||
81735 CTGTTTGGACCTTCTCATCAGTCCTCCATCAGAGAGGCGCCAGTCCCTGA 81784
|||||
44 uGlyValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysG 61
|||||
81785 AGGCCTCTGGGACCTTGGGTGCGATGGCTTCCTGCTCCAGCCCTGTG 81834
|||||
61 lYValGlyValGlnArgSerArgThrCysGlnLeuProThrValGln 77
|||||
81835 GGGTGGGGTGCAGCGAGGAGCGAATGT.....GAA 81869
|||||
78 LeuHisProSerLeuProLeuProProArgProArgHisProGluA1 94
|||||
81870 CTGACCCAGCCCTACCCCTCCCTCCCGAGCCCGCCAGACATCCAGAGC 81919
|||||
94 aLeuLeuProArgGlyGlnGlyProArgProGlnThrSerProGluThrL 111
|||||
81920 CCACGGGCGCGGGGCCAGGGCTCCAGACCCAGACTCCAGGGATCCCC 81969
|||||
111 euProLeuTyArgThrGlnSerArgGlyArgGlyGlyProLeuArgGly 127
|||||
81970 AGTCTCTGTATAGGCCACAGCCTCGGGAGAGGGTGGCCCTCTTCGAGCT 82019
|||||
128 ProAlaSerHisLeuGlyArgGluGluThrGlnGlu.IleArgAlaAlaA 144
|||||
82020 CCGTCTTCTCAAGTGGGAGAGAGAGAAACCCAGGAGCCTCAAGGGGCCA 82069
|||||
144 Tg..... 144
||
82070 GAGGTGAGGCTCTGAGGCAGAGCAGACAGGGGGTAAATGGAAAGGCCCT 82119
|||||
144 ..... 144
82120 TCAGCAGGCCAATCTTTCAGGAAGCAGACTACGGAGACAGGCTAAGAGAT 82169
|||||
145 .....ArgSerArgLeuArgAspProIleLysp 154
|||||
82170 GACTAAGTCTACATCTCTCTAGGTTCGGGTTCGAGACCCCATCAAGC 82219
|||||
154 roGlyMetPheGlyTyArgValPropheAlaLeuProLeuHisArg 170
|||||
82220 CAGGGATGTCGTTATGGAGAGTGCCTTTTGCTTGCTTCGACCGG 82269
|||||
171 AsnArgArgHisProArgSerPro.....ProArgSerGluLeuSerle 185
|||||
82270 AGCCCGAGGCATCTCTACACAGCCCGGCAACCCAGAAC..... 82308
|||||
185 uIleSerSerArgGlyGluGluAlaIleProSerProThrProArgAlaG 202
|||||
82309 ....TCTTCTACAGGGGAGGGATGTTCCATCTCAGCCTCCAGCAGACAG 82354
|||||
202 luProPheSerAlaAsnGlySerProGlnThrGluLeuPro....ProThr 217
|||||
82355 AGTGGCCCTCTGAAAAGACAGCGGCCCTCACAATGCAGCCCGCTGAACCCGCT 82404
|||||
218 GluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG1 234
|||||
82405 TCTCATCAGCTGAAACTCCA.....AGGCTCG 82433
|||||

```

234 uThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgH 251  
|||||  
82434 RACTGCTCAGACAGAAGTGCTCCGCCAAGTACTGCTGCTCT ..... 82476  
  
251 iHisProArgAlaGlnAlaSerGlyThrCluProProSerProHis 267  
82477 .....TCTTACACAGGACCCTGCACCTACCTCC 82506  
  
268 SerLeuGlyGluGlyGlyPheArgAlaSerProGlnProArgArgPr 284  
|||||  
82507 TCTTTGGAGACAGTAGGTCTTTCCAGGAGTCCCTTGGGCCAGGAATGCC 82556  
  
284 oSerSerGlnGly...TrpAlaSerProGlnValAlaGlyArgArg...P 299  
|||||  
82557 ACCTTCOCGGGAAGTTGGAGCATGCCAGGAGCAGACAGCCAGCACACC 82606  
  
299 toAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGly 315  
82607 CACCTCCCTTCTCTCTGTCCTCC...CGGAGCCAGCAGCAGGAGG 82650  
  
316 ProTrpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspPr 332  
|||||  
82651 CACTGG.....AGACTCCAGGTCTCCAGGTCCCCAGAT... 82686  
  
332 oGlnHisProGlyAlaTrpLeuProLeuSerAsnGlyProHisAlaAs 349  
82687 .....GCTGGCTGCCCCCTGACAAGAGC.....TCCA 82714  
  
349 erSerLeuTrpSerLeuPheAlaProSerSerProLeuProArgCysSer 365  
|||||  
82715 GCCACTCTGGAGCATCTTCTCCAGTATCCCTGCTCTCCAAACTGTCTCT 82764  
  
366 GlyGluSerGluGlnLeuArgAlaCysSerGln..... 376  
82765 GGGAAAGTGAACAGATGCGAGCTGCGAGCAGGAGGTGAGTTCTCTCC 82814  
  
376 ..... 376  
82815 CGGTTTGTGGCTAGGCTTTCACTGTTCCCTGACCTGGGTGCTCATCTT 82864  
  
376 ..... 376  
82865 GCCAGACTAGGACTGGAGATGGAGTTCTGACTTCAGGCGGAGCCCT 82914  
  
376 ..... 376  
82915 GATGTTCTCTGTGGAGCTGACCGACCGGTTCCCAACCCCTAGACTGGCT 82964  
  
376 ..... 376  
82965 GTGGCCACATCCACAGATTTCTCTCGAGCTCTGACTCTCTGTATCTGT 83014  
  
377 .....AlaProCysProGluGlnProAspProArgAlaLeuGlnCys 391  
83015 CTTCTCAGCCTTGGCCCCCTGAGCAGCAGACACCCAGGGCCCTACAGTGT 83064  
  
392 AlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTrpGluPr 408  
83065 GCGGCTTTGACTCCAGGAATTCATGGCCAGCTGTACCAGTGGGAGCC 83114  
  
408 oPheThrGlu.Val..... 412  
83115 CTTCCGAAAGTGGGTTTCTAGCCTGCOCOCACCCACCCACCCACCC 83164  
  
412 ..... 412  
83165 CCACCCCGAAGGACCGACTCAGGTTTGGTGGCCCTCTCTCCCTCTG 83214  
  
412 ..... 412  
83215 TTGTCTTACTTTTTCAGGTCAACACAGACACTTAGCCAGATCTAAGGAAG 83264

85065	CTATACAGCCANTGGGACTGTCCTCCAGTATACACCGTCTCCACGGGAAG	85111
564	luGlylsGlycyluSerLeuSerAlaGluGlyProThrThrGlnProVal	580
85115	AGGCAAGGGGAGAGICTGTGCAGAGAGGCCCTACCACTCAGCCGTGTG	85164
581	AspVal.Tyr	583
85165	GACGCTCTATGTGAGCTGGGCGAGGTGCTAGATCGTTAGGAGGAGGGAG	85214
583	.....	583
85215	GGCCCTTAGGCAGACAGCAGTGGTATCTAGCCTAGCGTATGCCCTCTTG	85264
584	.....MetIlePheGlnGluG1	589
85265	TTATCCCAAGCCATTGTGATCTTCCCTCCTCAGATGATCTTTCAGAGGA	85314
589	uAsnProGlyValPheTyGlnTyValIleSerSerProProIleLe	606
85315	CAACCAGAGTGTITTTATCAAGTATGTCATCTTCCCTCCTGCAGTCC	85364
606	euGluAsnProThrProGluProValProValProGlnLeuGlnPro.....	620
85365	TAGAGAGTCTTCCACAAAGCCTCCAGGCCCTCAGCCTCAGCCTCGGTAAA	85414
620	.....	620
85415	ACTAGAGTCCCCAGCACAGCAGAGAGGGGTGAGCGAGGACTGAGTCC	85464
621	.....Glu..Il	622
85465	AGGGTAGCAGGAGGACATCTAATCTGGTCCCTGGATTTCCTCGCAGAGAT	85514
622	eLeuArgValGluProProLeuAlaProAlaProArgProAlaArgThrP	639
85515	GCTGAGGGGGAGCCCTACVCCCTCAGCCCCCGCCGACATCGGGGCAC	85564
639	roGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProPro	655
85565	CAGGCACCTCCACGTCAGTGGGATTCGCCAGTGCCCTCCTCCGACT	85614
656	HisProArgThrProLeuGlySerProAlaAlaIleTyThrLysArgValGl	672
85615	CGTTCAGCAGACAGCCTGGGGTCTTCTCGTGACTGTGAACCAAGTGGG	85664
672	vhISerAlaCysSerAlaSerCysGlyLys.....	682
85665	GCACCTGAGTGTTCAGCATCTCTGTGGCAAGGTGAGGAGGACACCTCA	85714
682	.....	682
85715	GOCCAAATCCCTGTTCTCTATATATATCATACAGATTCTGTATCCTGGA	85764
682	.....	682
85765	CCACGGGTCTACATGGACCTATGCTGGGATTTCATTGAACCGGGCT	85814
683	.....GlyValT	685
85815	CTGAGCCTTCCTGCCTCCCTGGCTGCTGCTTATTCCTTCAGAGTGT	85864
685	rArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAsp	701
85865	GGCACCCCAATTTCTCTGTCATTTCCCTGTAGTCAGGAGAGGATTTGAT	85914
702	GluArgSerCysAlaAlaGlyAlaAlaArgProProAlaSerProGluProCy	718
85915	GAACAGAGCTGTGCTGGGGGCCAGACCCACAGCTTCCCTGAACCTGT	85964
718	shISGlyThrProCysProPro.....	725
85965	CCATAGACCCCGTGTCTCCATAGTGTAGTACGGCCACAGGGAGGGCTG	86014

[illegible]



725 ..... 725  
86015 GCCTGTCTTCTGCTCCCTCCCTCGGTTCTTACAGCACCTGGATCAGG 86064  
725 ..... 725  
86065 AAGATTAGATTAGAAAGTTTACACCCCTAGACAGTAAGAAACCTCG 86114  
725 ..... 725  
86115 CAGCTCTCTCCAGGCTAGAAATGAGATGCCAAGTGTGAATGACTCT 86164  
725 ..... 725  
86165 CTCCTATTGGGPTCCAGCTGTACACCCCTCCACATTTCCACCCAT 86214  
726 ..... TyrTrpGluAlaGlyGluTrpThrSerCys 735  
86215 CGTTCCTCTCCCTACCCAGCTGGGAGCTGGGAGTGGGAGTGGACATCC 86264  
735 sSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArg 752  
86265 CAGCCGATCTGTGCGCTGGCCAGCCAGCCAGCTACTCTGCAGAC 86314  
752 InGluPheGlyGlyGlySerSerValProGluArgCysGlyHis 768  
86315 AGGAATTCGAGTGTGTGCTCTCGTACCTCCAGAGCGCTGTGGACAT 86364  
769 LeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHis 785  
86365 CTCCTCCGCGCCCAACATCACCCAGCTTGTACAGCTGCACCTCTGTGCCA 86414  
785 sTrpGluValGlySerProTrp.Ser..... 793  
86415 CTGGAGATTAGCTCCCTCCCTGGAGCCAGGTGAGGAGGACCTTGAGAAT 86464  
793 ..... 793  
86465 AGCTCCCAAGAATTTCCACCTTTCTTCTTCTCATCAGAACTTCCA 86514  
793 ..... 793  
86515 AACAGACCTCTATGGAACCTAGTATATATATATATATATATATATAT 86564  
793 ..... 793  
86565 GCCTGGAGGCCCTGTGTACCTCCCTCCCTGTGTGAACCTAGTCTGT 86614  
793 ..... 793  
86615 AGAAGCAAGCTGGAAGATGGCTACGCCAGCTCGGGTCTCTCATCT 86664  
794 GlnCysSerValArgCysGlyArgGlyGlnArgSerArgGlnValArgC 810  
86665 TCAGTGTCTGTGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 86714  
810 YsValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAlaSerGly 826  
86715 GTGTGGCACTAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 86764  
827 ProProGlnProProSerArgGluAlaCysAspMetGlyProCysThrTh 843  
86765 CCCCCCGCGCCGCCAGCAGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGT 86814  
843 rAlaTrpPheHisSerAspTrpSerSerLysValSerPro 856  
86815 AGCTGTGTCTTACAGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 86854

seq\_name: gb\_htg:AC093317

seq\_documentation\_block:

LOCUS

215810 bp DNA linear

HTG 29-JAN-2002

## DEFINITION

Mus musculus clone rp23-218k6 strain C57BL/6J, WORKING DRAFT  
SEQUENCE, 10 unordered pieces.

## ACCESSION

AC093317

## VERSION

AC093317.16 GI:18390243

## KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

## SOURCE

house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 215810)

## AUTHORS

Jia H., Zhang P., Lin S. and Roe, B.A.

## TITLE

Mus musculus BAC Clone rp23-218k6

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 215810)

## AUTHORS

Jia H., Zhang P., Lin S. and Roe, B.A.

## TITLE

Direct Submission

## JOURNAL

Submitted (17-AUG-2001) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

## COMMENT

On Jan 29, 2002 this sequence version replaced gi:18139414.

## -----

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

## -----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.\* 1 2528: contig of 2528 bp in length  
\* 2529 2628: gap of unknown length  
\* 2529 8257: contig of 5629 bp in length  
\* 8258 8357: gap of unknown length  
\* 8358 14071: contig of 5714 bp in length  
\* 14072 14171: gap of unknown length  
\* 14172 28167: contig of 13996 bp in length  
\* 28168 28267: gap of unknown length  
\* 28268 41090: contig of 12823 bp in length  
\* 41091 41190: gap of unknown length  
\* 41191 58658: contig of 17468 bp in length  
\* 58659 58759: gap of unknown length  
\* 58759 83163: contig of 24405 bp in length  
\* 83164 83263: gap of unknown length  
\* 83264 107370: contig of 24107 bp in length  
\* 107371 107470: gap of unknown length  
\* 107471 137048: contig of 29578 bp in length  
\* 137049 137148: gap of unknown length  
\* 137149 215810: contig of 78662 bp in length.

## FEATURES

## source

1..215810  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="rp23-218k6"BASE COUNT 59907 a 50405 c 47367 g 57018 t 1113 others  
ORIGIN

## alignment\_scores:

Quality: 2651.50 Length: 1751  
Ratio: 3.804 Gaps: 29  
Percent Similarity: 39.806 Percent identity: 35.294

## alignment\_block:

US-10-041-770-2 x AC093317 ..

Align seg 1/1 to: AC093317 from: 1 to: 215810

```

28 LeuSerGlyHisSerLeuGlnThrProThrGluGluGlyGlnGlyProG1 44
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109815 CTGTTGGACCTTCTCATCAGTCGCCATCAGAGGAGGGCCAGTCCCTGA 109864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 uGlyValTTPGlyProTTPValG1LTPAlaSerCysSerGlnProCysG 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109865 AGCCCTCTGGGACCTTGGGTGCGATGGGCTTCCCTGCTCCAGCCCTG 109914
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 lyVal.GlyValGlnArgSerArgThrCysGlnLeuProThrValG1 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109915 GGGTGGGGGTGCAGCGCANGAGCGCAACATGT.....GA 109949
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77 nLeuHisProSerLeuProLeuProProArgProProArgHisProGluA 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109950 ACTGCACCCAGCCCTACCCCTCCCTCCCGACCCCAAGACATCCAGAG 109999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 laLeuLeuProArgGlyGlnGlyProArgProGlnThrSerProGluThr 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110000 CCCACCGCCCGGGCCAGGCTCCAGACCCAGACTCCACGGGATCC 110049
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 LeuProLeuTyArgThrGlnSerArgGlyArgGlyGlyProLeuArgG1 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110050 CAGTCTCTATAGGCCACAGCCCTCGGGAGGGGTGGCCCTCTTCGAGC 110099
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
127 yProAlaSerHisLeuGlyArgGluGluThrGlnGlu.IleArgAlaAla 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110100 TCCTGCTTCTCAGTGGGAGAGAGGAAACCCAGAGCCTCAGGGGCC 110149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 Arg..... 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110150 AGAGGTGAGGCTCTGAGGCGAGAGAGAGAGGGGTAATGAAAGGCG 110199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 ..... 144
110200 TTCAGAGGCCAATCTTTCAGGAGCAGACTACGGAGACAGGCTAAGAGA 110249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ..... 145
110250 TGACTAAGTCTACATCCTCCTCCTAGGTTCCGGGTTCGAGACCCCATCAAG 110299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 ProGlyMetPheGlyTyArgValProPheAlaLeuProLeuHisAr 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110300 CCAGGAGATGTCGGTATGGAGAGAGTCCCTTTTCCTTCCTCTGCACCG 110349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 gAnArgArgHisProArgSerPro.....ProArgSerGluLeuSerL 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110350 GAGCGCAGCATCCTCAGACAGCCCGGCAACCCCAAGAAC..... 110389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 euLeSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110390 .....TCTTCTCAGGGGAGGAGTGGTTCATCTCAGCCTCCAGCACA 110434
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
202 GluProPheSerAlaAsnGlySerProGlnThrGluLeuPro...ProTh 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110435 GAGTGGCCCTCGAAAAGCAGCGCCCTCAGATGCAGCCCCCTGAACCCG 110484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110485 TTCTCACTCAGTGAACCTCCA.....AGGTCGTG 110513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 luThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110514 GAAGTGTCTACAGAGAGTGTCTCCAGAGACTAGTTCGTCTCT 110557
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 HisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110558 .....TCTACACAGAGGACCCCTGCACTACCTC 110586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 sSerLeuGlyGluGlyPhePheArgAlaSerProGlnProArgArgP 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110587 CTCCTTTGGAGACAGTAGTGTCTTCAGGATCCCTTGGGCCACGATGC 110636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
284 roSerSerGlnGly...TrpAlaSerProGlnValAlaGlyArgArg... 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

111495 CTTGTGGCTATGTCACCTCCACACCAAGACAAACCTTCATGAGCTATGAA 111544  
412 ..... 412  
111545 CTCATGATCACAGTGTGAGAAACCTTTTGAGGAATGGTAGGCTTGGCT 111594  
412 ..... 412  
111595 CAGTGGGTAAAGAAAGCTGTCTCAAGCCTPAGTGACCTGAGTTGATCTCC 111644  
412 ..... 412  
111645 AGAATCCAGAACCCGCTCTCTCAAGATGACCTCTTCTCCACCTGAGC 111694  
412 ..... 412  
111695 ACCCAATCTTATACAAAGGAAACAAACAAACAAATCTTTGATGSC 111744  
412 ..... 412  
111745 CAGCTGCCAAACTTCCCTTTTATAGGAGCAGAAATGGAAGAACAGTC 111794  
412 ..... 412  
111795 AAATCTCTTTTTTTTTTTTTTTTTTTTTTTTTTCTGAGACGGGTTT 111844  
412 ..... 412  
111845 CTCTGTGCAGTCTGCTGCTCTGGAACCTCTGTACACAGGCTGGC 111894  
412 ..... 412  
111895 CTCAAACTCAGAAATCTGCTTGCTCTGCTCCCAATGCTGGATTAA 111944  
412 ..... 412  
111945 GGTGTGCACCAACCGCTGGCATCATCTTAATCTTAACCTGCAGACCA 111994  
412 ..... 412  
111995 CATAGCATAACAAAAACATAAGACACAAATCGTAAGCGTGGAAAGTTG 112044  
412 ..... 412  
112045 CCGCTCCCGCTTCCACACAGCTTCCCGATGCCCGCATATGTAAAGC 112094  
412 ..... 412  
112095 TGTCAGGCTCTTGGGGGGGGTGGCTTTCTCGAGGGAACCCAGTTAA 112144  
413 ..... GlnGlySerGlnArgCysGluLeuA 421  
112145 CTAGCCTGTGTCTCTCTCTTTAGTTACGGGCTCCACGCGTGTGAAGTGA 112194  
421 snCysArgProArgGlyPheArgPheTyrValArgHisThrGluLysVal 437  
112195 ACTCCGCGCGCGTGGTTCGGATTCATATCCGACACACTGAAAGGTTG 112244  
438 GlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysValAlaG 454  
112245 CAGGATGGAACCTGTGTACGCTGGATCCCTAGACATCTGTGTGGCTGG 112294  
454 YArgCysLeu...Ser..... 458  
112295 ACGTGTGCTGTGAGGAGGCGAGGCTGTGGTTGTACATGTGCACCTA 112344  
458 ..... 458  
112345 CACACACACACACACACACACACACACACACACACTCCCTTAATTG 112394  
458 ..... 458  
112395 CCACAAGAGTGAAACACAGACTGAGGTGGCCAAAGGGAAGTTTCTGC 112444

458 ..... 458  
112445 TGTCAGTAAAGATGGTGTGTAGACAGATCCACAGACTTGCTAAAG 112494  
458 ..... 458  
112495 CTCCTCCAGACATAGTGGGAGGAGGAGGTCAGAAATACTCTGACTTG 112544  
458 ..... 458  
112545 AAGCACAAGAGATGTGGCAGCAGGAGGTGAGCTGAGGGCCCTGGGTGGAG 112594  
458 ..... 458  
112595 CGGGCCAGGTGTGAGGATCCCTGGGCTGTGTGTGGCCCTGCTCACC 112644  
459 ..... ProGlyCysAspGlyIleLeuGlySerGlyArgA 470  
112645 TCACCTCTCCACAGAGCCCTGGCTGTGATGGGTCTTGGCTCTGGCAGGC 112694  
470 rgProAspGlyCysGlyValCysGlyGlyAspAspSerThrCysArgLeu 486  
112695 GTCCGATGGCTGGAGTCTGCGGGGTGATGTTCTACCTGCCGCTG 112744  
487 ValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLysII 503  
112745 GTTTCGGAAATCTCACTGACCGAGGGGCGCTTGGGCTATCAGAAAT 112794  
503 eleuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgP 520  
112795 CTTGTGGATCCCTGCCGGGCTCCACCTTCACATTTCCAGTTGGNC 112844  
520 roSerSerAsnTyrLeu..... 525  
112845 CCAGTTCCAATTACCTCGGTAAGAAAGAGGACCTCGGTCTCTATGCCCGAG 112894  
525 ..... 525  
112895 TGTATCTTTTCTTTGATCTGCTTACACACCTCTTACGCGGAACG 112944  
525 ..... 525  
112945 AAGCAAAACAGGAAGGCTGGAGCCTTCTCATAAACCGATGGAGG 112994  
525 ..... 525  
112995 TGAGGAGGTGGATGTCAGGTCTCTGGGCACAAAGGCTTCAGAGAGA 113044  
526 ..... AlaLeuArgGlyPyr 530  
113045 GTACCACTATCACTCAGCTCAGACATCTGCTTTTCAGCAGCTCCGAGGGC 113094  
530 oglyGlyArgSerIleIleAsnGlyAsnTrpAlaValaspProProGlyS 547  
113095 TGGGGCGGCTCCATTATCAATGGGAAGTGGGCTGTGGATCTCCAGGT 113144  
547 erTyrArgAlaGlyGlyThrValPheArgTyrAsnArgProProArgGlu 563  
113145 CCTATACAGCCATTGGGACTGTCTCCAGTATACCGTCTCCACGGGA 113194  
564 GluglyLysGlyGluSerLeuSerAlaGluGlyProThrGlnProva 580  
113195 GAAGCAAGGGGAGAGTCTGTGACAGAAAGGCCCTTACCACTCAGCCTGT 113244  
580 lAspVal.Tyr..... 583  
113245 GGACGCTTATGTAGCTGGGCGAGGTGTAGATCGTTAGGAGGGAGGA 113294  
583 ..... 583  
113295 GGGCCCTTAGGCAGACAGCAGTGTGGATCTAGCCTAGCGTATGCCCTCTT 113344

[illegible]

seq\_documentation\_block:  
 LOCUS AC053497 207815 bp DNA linear HTG 13-AUG-2000  
 DEFINITION Homo sapiens chromosome 1 clone RP11-243G22, WORKING DRAFT  
 SEQUENCE. 37 unordered pieces.

VERSION AC053497.4 GI:9797901  
KEYWORDS HTG: HTGS PHASE1: HTGS DRAFT.

ORGANISM	Homo sapiens
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

REFERENCE 1 (bases 1 to 207815)  
MAMMALIA; EULHERIA; FILICES;

**TITLE** The sequence of Homo sapiens

—

AUTHORS  
TITLE  
JOURNAL  
COMMENT

Waterston,R.H.  
Direct Submission  
Submitted (16-APR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Aug 12, 2000 this sequence version replaced gi:8439979.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_NH0243G22  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 192323 bases at least Q40  
Consensus quality: 196320 bases at least Q30  
Consensus quality: 198661 bases at least Q20  
Insert size: 204000; agarose-fp  
Insert size: 204215; sum-of-contigs  
Quality coverage: 6.36 in Q20 bases; agarose-fp  
Quality coverage: 6.31 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1
* 1381: contig of 1381 bp in length
* 1481: gap of unknown length
* 1482: contig of 1121 bp in length
* 2602: gap of unknown length
* 2603: contig of 1311 bp in length
* 2703: contig of 1311 bp in length
* 4014: gap of unknown length
* 5331: contig of 1218 bp in length
* 5332: gap of unknown length
* 5431: gap of unknown length
* 5432: contig of 1191 bp in length
* 6623: gap of unknown length
* 6723: gap of unknown length
* 8166: contig of 1444 bp in length
* 8167: gap of unknown length
* 8267: gap of unknown length
* 9674: contig of 1408 bp in length
* 9675: gap of unknown length
* 9775: contig of 1169 bp in length
* 10943: gap of unknown length
* 10944: contig of 1155 bp in length
* 11044: gap of unknown length
* 12198: gap of unknown length
* 12199: contig of 1895 bp in length
* 12299: gap of unknown length
* 14194: gap of unknown length
* 14294: contig of 1424 bp in length
* 15717: gap of unknown length
* 15817: contig of 1118 bp in length
* 15818: gap of unknown length
* 16936: gap of unknown length
* 17035: contig of 1340 bp in length
* 18375: gap of unknown length
* 18376: gap of unknown length
* 18475: gap of unknown length
* 18476: contig of 1487 bp in length
* 19862: gap of unknown length
* 20062: contig of 1107 bp in length
* 21169: gap of unknown length
* 21269: gap of unknown length
* 22452: contig of 1183 bp in length
* 22552: gap of unknown length
* 24582: contig of 2030 bp in length
* 24583: gap of unknown length
* 26383: contig of 1701 bp in length
* 26384: gap of unknown length
* 27170: contig of 1227 bp in length
* 27810: gap of unknown length
* 27711

```

```

* 27811
* 29195: contig of 1385 bp in length
* 29196: gap of unknown length
* 29295: contig of 1505 bp in length
* 30800: gap of unknown length
* 30801: contig of 1655 bp in length
* 32555: gap of unknown length
* 32556: contig of 1998 bp in length
* 34553: gap of unknown length
* 34554: gap of unknown length
* 34754: contig of 2369 bp in length
* 37122: gap of unknown length
* 37123: contig of 2284 bp in length
* 39506: gap of unknown length
* 39507: gap of unknown length
* 39607: contig of 1614 bp in length
* 41320: gap of unknown length
* 41321: contig of 1638 bp in length
* 42978: gap of unknown length
* 43079: contig of 2695 bp in length
* 45774: gap of unknown length
* 45873: contig of 9183 bp in length
* 55056: gap of unknown length
* 55156: gap of unknown length
* 55057: contig of 11903 bp in length
* 55157: gap of unknown length
* 67060: contig of 13101 bp in length
* 67160: gap of unknown length
* 80261: contig of 16168 bp in length
* 80360: gap of unknown length
* 96528: gap of unknown length
* 96529: contig of 17901 bp in length
* 114529: gap of unknown length
* 114530: contig of 19044 bp in length
* 133673: gap of unknown length
* 133674: contig of 20018 bp in length
* 133773: gap of unknown length
* 153791: contig of 20018 bp in length
* 153792: gap of unknown length
* 153891: contig of 22188 bp in length
* 153892: gap of unknown length
* 176079: gap of unknown length
* 176179: contig of 31636 bp in length
* 176080: gap of unknown length
* 207815: contig of 31636 bp in length
* 176180: gap of unknown length

```

## FEATURES

```

source
1. .207815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-243C22"
misc_feature 1..1381
/note="assembly_name:Contig13"
misc_feature 1482..2602
/note="assembly_name:Contig15"
misc_feature 2703..4013
/note="assembly_name:Contig21"
misc_feature 4114..5331
/note="assembly_name:Contig25"
misc_feature 5432..6622
/note="assembly_name:Contig33"
misc_feature 6723..8166
/note="assembly_name:Contig37"
misc_feature 8267..9674
/note="assembly_name:Contig38"
misc_feature 9775..10943
/note="assembly_name:Contig40"
misc_feature 11044..12198
/note="assembly_name:Contig42"
misc_feature 12299..14193
/note="assembly_name:Contig45"
misc_feature 14294..15717
/note="assembly_name:Contig47"
misc_feature 15818..16935
/note="assembly_name:Contig50"
misc_feature 17036..18375
/note="assembly_name:Contig51"
misc_feature 18476..19962
/note="assembly_name:Contig52"
misc_feature 20063..21169
/note="assembly_name:Contig53"
misc_feature 21270..22452
/note="assembly_name:Contig54"

```



529 ..... 529  
132519 AGGTTCTGGTGGGAGCTTCTATAGGCTAAGGACACGGTGTGGGAGGGA 132470  
529 ..... 529  
132469 AGGTATTATACCCCTGGGAATTTCCGGATCTCTACCTCTGACCCGCGAGCA 132420  
530 .....ProGlylyArgSerIleIleAsnGlyAsnTrpAlaValasp 543  
132419 CTTCTGGCCCTGGGGCCGGTCCATCATCAATGGGAACCTGGGCTGTGGAT 132370  
544 ProProGlySerTyrArgAlaGlyVlyThrValPheArgTyrAsnArgPr 560  
132369 CCCCCTGGGTCTTACAGGGCCGGGGGACGTC.TTTCGATATACCGTCC 132321  
560 oProArgGluGluGlyLysGlyGlySerLeuSerAlaGluGlyProThr 577  
132320 TCCAGGGAGGAGGGCAAGGGGAGAGTCTGTGGCTGAAGGCCCAACA 132271  
577 hrGlnProValaspValtyr ..... 583  
132270 CCCAGCCTGTGGATGTCTATGTAGCCCTGGGGCCAGGGGAGCTGAATGCT 132221  
583 ..... 583  
132220 GGGAGGGAGGCTGTTCCTCTGGGACAGCTGTGGTTGTCAAGATGGGA 132171  
583 ..... 583  
132170 GAGAGAGCTGGTGGCATCCTCTTGGCCACCCCAATATCATATATCT 132121  
584 .....MetilePheGlnGluAsnProGlyValPheTyr 595  
132120 TCTCTTCTCCCGATGATCTTTCAGGAGGAAACCCAGGGCTTTTAT 132071  
596 GlnTyrValIleSerProProProlleLeuGluAsnProThrProGl 612  
132070 CAGTATGTCTCTTCACTCTCTCAATCTTGAGAACCCCAACCCAG 132021  
612 uProProValProGlnLeuInPro ..... 620  
132020 GCCCCCTGTCCCCAGCTTTCAGCCGGGTAAAGACTGTGACCCCTGCACITG 131971  
620 ..... 620  
131970 GAAGGAGGAGGAGAGGCTTCAGGGCTGGCTCGGGGAGTGGGGTGGCAT 131921  
621 .....GluIleLeuArgValGluProProle 629  
131920 CTGATTCGCTGCTCCCTGACAGAGATCTGAGGGTGGAGCC.CCACT 131872  
629 uAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnArgGlnV 646  
131871 TGTCCGGACCCCGCCAGCCCGGACCCAGGACCCCTCCAGGCTCAGG 131822  
646 alArgIleProGlnMetProAlaProHisProArgThrProLeuGly 662  
131821 TGGGGATCCCGCAGATGCCCGCCCGCCCGCCATCCAGGACCCCTGGGG 131772  
663 SerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSe 679  
131771 TCTCCAGCTCGGTACTGGAAACAGTGGACACTCTGCATGTCTCAGCGTC 131722  
679 rCysGlyLysGly ..... 683  
131721 CTGGGAAAGGTGAGACATCACAGTGGCTTCCCGCATCTCGGTCCAAA 131672  
683 ..... 683  
131671 CCCCCCAACTGACACTCCCGCATCTCTGGATTGTGGGGCCAGCCCGACCC 131622  
683 ..... 683

131621 TGGCTCTGGGACACCACTAGCTTGGGCTTAGCTCTAGCTCTCCCTCCCTGGC 131572  
684 .....Val..TyrArgProIlePheLeuCysIleSe 693  
131571 TGCCTTCTCACCCACTCAGGTGTCTGGCGCCCATTTCTCTCTGCATCTC 131522  
693 rArgGluSerGlyGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
131521 CCGTGAGTCGGGAGAGAACTGGATGAACGACAGCTGTGCCCGGGTGCCA 131472  
710 rGProProAlaSerProGluProCysHisGlyThrProCysProPro... 725  
131471 GGCCCCAGCCTCCCTGTGAACCTGCGACGGCACCCCATGCCCCCATAG 131422  
725 ..... 725  
131421 TGAGTATGGGGAGCCCAACGGGAGGGTTAGGTACTTGGAAACACAGCAG 131372  
725 ..... 725  
131371 TTGCCCCCAGAATCTTACCTGAACCTCTGCACCTCAACGCAACATCTCTCTGG 131322  
725 ..... 725  
131321 CTGCAGAGAGACAACATTAGATTAGAAAGCAGACATTCGGTAGGCAGCA 131272  
725 ..... 725  
131271 AGAAGCCATGGCAGGCTGAGCCCCCAGGACCCAGACTCCAACGCCA 131222  
725 ..... 725  
131221 AACGTGCCACTCTCTCTGAGGCCCCAGCTCCAGTGTGTCTTCCATCT 131172  
725 ..... 725  
131171 CTGCTGCCAGGGAGCCCCACACATCTCATCTATGTCTCCGCTCTTCCC 131122  
726 .....TyrTrpGluAlaGlyGluTrpThrSerCysSe 736  
131121 TGCCTGTGTGTGCTGCAGCTGGGAGGCTGGCGAGTGGACATCTCTCAG 131072  
736 rArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnG 753  
131071 CCGTCTCTGTGGCCCCGCGCACCCAGCACCCGACCTGCAGTGCAGTCCGGCAGG 131022  
753 luPheGlyGlyGlySerSerValProProGluArgCysGlyHisLeu 769  
131021 AATTTGGGGGGTGGCTCTCTCGGTGCCCGGAGCGCTGTGGACATCTC 130972  
770 ProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisTr 786  
130971 CCGCGGCCAACATCACCCAGTCTTGCCAGCTGCGCTCTGTGGCCATTG 130922  
786 pGluValGlySerProTrp.Ser ..... 793  
130921 GGAAGTTGGCTCTCTCTGGAGCCAGGTGAGTTTGCACGCAAGGAGGTG 130872  
793 ..... 793  
130871 CTGGGAGGGGAATGGGCACAGGTGAACAACAGCAGTGGTTTTTCAACTT 130822  
793 ..... 793  
130821 CTTTTCATCAGCAAACTATTATACAAACAGGTTTTCATATGGACCCCC 130772  
793 ..... 793  
130771 AATATAGAAGTACAGATAAGGAGTGAACAGGAGTGGGGGCCAGGCC 130722  
793 ..... 793





```

vector_side:left"
45874. .55056
/note="assembly_name:Contig70"
55157. .67059
/note="assembly_name:Contig71"
67160. .80260
/note="assembly_name:Contig72"
80361. .96528
/note="assembly_name:Contig73"
96629. .114529
/note="assembly_name:Contig74"
114630. .133673
/note="assembly_name:Contig75"
133774. .153791

alignment_scores:
  Quality: 2014.50      Length: 834
  Ratio: 4.304          Gaps: 14
  Percent Similarity: 56.115  Percent Identity: 52.998

alignment_block:
US-10-041-770-2 x AC053497 ..

Align seg 1/1 to: AC053497 from: 1 to: 207815

7 ArgProTrpLeuYrLeuLeuLeuSerLeuProGlnLeuCysLe 23
|||||
64556 AGGCCCTGGCTGATCTGCTGCTGCTTGTGTCCTCCCTCAGCTCTGCTT 64605
23 uaspGlnGlu..... 26
|||||
64606 GGATCAGGAGGTGAGTTCTGGACAAGTGAGCAGCTGCAGCTGCTGCCA 64655
26 ..... 26
64656 CCCTTTCATCTCCCTTAGGCTCCACCCCATCTCTCCAGGCCACGCTCTC 64705
26 ..... 26
64706 CATTCGCCACAGCCACCCACCTCCCTGCCAAGTCCACAGTTCCTGCC 64755
26 ..... 26
64756 TCATACACCAAGAGCGGAGGTGTAGTTCTCCCTCTGCTGCTGAATGTG 64805
26 ..... 26
64806 ACCTTGGACTGTAGCAGCTCCGTGAGCCTCAGTGTGGTCTACATGGAC 64855
26 ..... 26
64856 ACCTGGACAGATTCCTCTAATCTCTCAATCTCTGTCACCTTTTCTC 64905
27 ..... ValLeuSerGlyHisSerLeuGlnThrProThrGluGluGlyG 41
|||||
64906 TATATAGGTGTGTCCGAGACACTCTCTTCAGACACTACAGAGGAGGGCC 64955
41 lnGlyProGluGlyValTrpGlyProTrpValGlnTrpAlaSerCysSer 57
|||||
64956 AGGCCCCCGAAGGTGTCGGGACCTTGGTCCAGTGGCCCTTGTCTCC 65005
58 GlnProCysGlyValGlyValGlnArgArgSerArgThrCysGlnLeuPr 74
|||||
65006 CAGCCCTCGGGGTGGGGGTGCAGCGCAGGAGCGGACATGTCAGCTCCC 65055
74 oThrValGlnLeuHisProSerLeuProLeuProProArgProProArgH 91
|||||
65056 TACAGTCAGCTTCACCCGAGTCTGCCCTCTCCCTCCCGGCCCCCAAGAC 65105
91 isProGluAlaLeuLeuProArgGlyGlnGlyProArgProGlnThrSer 107
|||||
65106 ATCCAGAGCCCTCTCTCCCGGGGCCAGGGTCCAGAGCCCGAGACTTCT 65155

```

```

108 ProGluThrLeuProLeuTyrArgThrGlnSerArgGlyArgGlyGlyPro 124
|||||
65156 CCAGAAACCCCTCCCTTGTCACAGGACACAGTCTCGGGAAGGGGTGGCC 65205
124 OLeuArgGlyProAlaSerHisLeuGlyArgGluGlnThrGlnGluIleA 141
|||||
65206 ACTTCAGGTCCCGCTTCCACCTAGGGAGAGAGAGAGAGAGAGAGATTC 65255
141 rGAlaAlaArg..... 144
|||||
65256 GAGCGCCAGAGGTGAGAGGCTGGGTGGAAGAGGTGGGCTTGGCAA 65305
144 ..... 144
65306 GGTGGGGCTTAGCATGAAGAAAGGGAGCACGGGTGGCTTCAGAGGC 65355
144 ..... 144
65356 TTGCTTTGGTCCAGGAAGTAACCTGGGGAACAGGGTCAGAGTGGTC 65405
145 .....Arg.SerArgLeuArgAs 150
|||||
65406 AGAGCTGTGCCCCACATCTGAACACCTTCTCAGGTCCCGCTTCGAGA 65455
150 pProIleLysProGlyMetPheGlyTyrGlyArgValProPheAlaLeuP 167
|||||
65456 CCCATCAGCCAGGAATGTTCGGTTATGGGAGAGTCCCTTTGCATTGC 65505
167 rOLeuHisArgAsnArgHisProArgSerProProArgSerGluLeu 183
|||||
65506 CACTGCACCGGAACCCAGCACCTCGGAGCCACCACCCAGATCTGAGCTG 65555
184 SerLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProAr 200
|||||
65556 TCCTGATCTCTTCTAGAGGGGAAGAGGTATTCCCTGCCCTACTCCAAG 65605
200 gAlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProT 217
|||||
65606 AGCAGACCCATTCTCCGAAACGGCAGCCGCCCAAACTGAGCTCCCTCCA 65655
217 hrGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerPro 233
|||||
65656 CAGAACTGTGTCTCCACACCCCTTCCGCCCAAGCAACCACTCTAAGCCCT 65705
234 GluThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuAr 250
|||||
65706 GAACCTGCTCAGACAGAGGTGGCCCCAGAACCCAGGCTGC.CCCCTACG 65754
250 gHisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrH 267
|||||
65755 GCATCACCAGAGCCAGGCTGTGCACAGAGCCGCCCTCACCACGC 65804
267 lSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArgArg 283
|||||
65805 ACTCCTTAGGAAAGTGGCTTCTTCGTGCATCCCTCAGCCACGAAAG 65854
284 ProSerSerGlnGlyTyrAlaSerProGlnValAlaGlyArgArgProAs 300
|||||
65855 CCAAGTTCACAGGTTGGCCAGTCCCCAGGTAGCAGGGAGACGCCCTGA 65904
300 pProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProT 317
|||||
65905 TCCCTTTCCTTCGCTCCCTCGGGGCCAGGCCAGCA.GGGCAAGGGCCTT 65953
317 rpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspProGln 333
|||||
65954 GGGAAACGGGGGACTCCTCACGGGCCCGCCCTGGAGCCTGACCCCTCAG 66003
334 HisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSerSe 350
|||||
66004 CACCCGGCGCTGGCTGCCCTGCTGAGCAACGGCCCAATGCCAGCTC 66053
350 rLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyG 367
|||||
66054 CCTGTGAGCCTCTTTGCTCCAGTAGCCCTATTCCCAAGATGTTCTGGG 66103
367 luSerGluGlnLeuArgAlaCysSerGln..Ala..... 377
|||||
66104 AGAGTGAACACTAAGAGCCTGCAGCCCAAGCGGTGAGTCTCCTCGGGCCT 66153
377 ..... 377
66154 CCCCTCCCAACCCCGACCTCCAGTGTGGCTTCCCTGCCCTGAAGCTGGGT 66203
377 ..... 377
66204 CTTCAGCTTCGCTTGGCACCTGAGGGAGAGGCCCTTGGCATCTGACCA 66253
377 ..... 377
66254 CCTCAGGCGAGGGGTCTTTGGAGCTCTTCCGCTGACCTGAGCCTCCCACT 66303
377 ..... 377
66304 GCTCCCCAGGTTACGCCCTGCCCTACCTCATTTGCTCCCCAGCTCTG 66353
378 .....ProCysProGluGlnProAspPr 386
|||||
66354 ACTCCTTTGTACCCCTCACCAGCCCTGCCCTCCCTGAGCAGCCAGACCC 66403
386 oArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGlnL 403
|||||
66404 CCGGCCCTGTCAGTCCGAGCCTTTAACTCCAGGAATTCATGGGCCACG 66453
403 euTyrGlnTrpGluProPheThrGluValGlnGlySerGlnArgCysGlu 419
|||||
66454 TGTATCAGTGGGAGCCCTCACTGAA.....GGTGAG 66485
420 LeuAsnCysArgProArgGlyPheArgPheTyrValArgHisThrGluLy 436
|||||
66486 GTTCTTGTCTACCCCTGGGC..... 66506
436 sValGlnAspGlyThrLeuCysGlnProGlyAlaProAspIleCysVala 453
|||||
66507 .....AGTGGTGGCTTGGAAATTTGGGATGA 66531
453 laGlyArg.....CysLeuSerProGlyCysAspGlyIle..... 464
|||||
66532 AGGGAGAGGAGATACCTGCTTACTCCAGCCCTGAATGACTTCCAGCC 66581
465 .....LeuGlySerGlyAr 469
|||||
66582 CTCTGCTTCCCTGGCCATGCCTTCTTCTTCTCCCTGGGCTGGGTCA 66631
469 gArgProAspGlyCysGlyValCysGlyGlyAspAspSerThrCysArgL 486
|||||
66632 G.....GAGACAGCACAGG 66645
486 euValSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnLys 502
|||||
66646 TGGTGAAGTGTCTGCAGAGGGTGGGGTGGG..... 66677
503 ileLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuAr 519
|||||
66678 .....AG 66679
519 gProSerSerAspTyrLeuAlaLeuArgGly.ProGlyGlyArgSerIle 535
|||||
66680 GAGAGGTGTGGAGACACGCTTTTGTCCGACTCCCTCGGAAGGCCATC 66729
536 ileAsnGlyAsnTrpAlaVal.....AspPr 544
|||||
66730 ACTGGGGTCAAGTGGCTGTGCACACAAGAGGCCATGTTGGGAGAGATCC 66779
544 oProGlySerTyrAlaGlyGlyThrValPheArgTyrAsnArgProp 561
|||||

```

```

66780 TGTCCAGCGTGACAGCCCGGTGGGGGTGTG.....C 66811
561 rGArgGluGluGlyGlySerLeuSerAlaGluGlyProThr 576
||||| ||| |||:||||: |||: |||: |||:
66812 CAGGCATCTGGGCACTGCGTGGTGTGCTCCAGGTACCATCC 66858

seq_name: gb_pr:AF217974
seq_documentation_block:
LOCUS AF217974 2070 bp mRNA linear PRI 01-OCT-2000
DEFINITION Homo sapiens clone PP1396 unknown mRNA.
ACCESSION AF217974
VERSION AF217974.1 GI:10441878
KEYWORDS FLI_CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2070)
Gu.J.R., Wan.D.F., Zhao.X.T., Zhou.X.M., Jiang.H.Q., Zhang.P.P.,
Qin.W.X., Huang.Y., Qiu.X.K., Qian.L.F., He.L.P., Li.H.N., Yu.Y.,
Yu.J. and Han.L.H.
Novel Human cDNA clones with function of inhibiting cancer cell
growth
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2070)
Gu.J.R., Wan.D.F., Zhao.X.T., Zhou.X.M., Jiang.H.Q., Zhang.P.P.,
Qin.W.X., Huang.Y., Qiu.X.K., Qian.L.F., He.L.P., Li.H.N., Yu.Y.,
Yu.J. and Han.L.H.
Direct Submission
TITLE Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related
Genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-Tu Road, Shanghai
200032, People's Republic of China
LOCATION/Qualifiers
FEATURES
Source
1..2070
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pp1396"
602..1315
/codon_start=1
/product="unknown"
/protein_id="AAG17217.1"
/db_xref="GI:10441878"
/translation="MGCTTAFHSDWSSKCSAECTGIORRSVVCLGSGNALPGGOG
EAGTGQSCPTGSRPDMRACSLGCEKTRWRTGFWGBCSSCGSGTORRDILCVS
LKGTFENVTSFNSHPALQCGQACQDRFWSTPWSPCRSQGGTQTRREVOC
LSTNQTLSRCPQLRPSRKRPCNSQPCSRPDDQCKDSSPHCPVLVYVQARLCVPPYT
ATCCRSCHVLESPQDPS"
BASE COUNT 409 a 670 c 585 g 406 t
ORIGIN

alignment_scores:
Quality: 1231.00 Length: 220
Ratio: 5.673 Gaps: 0
Percent Similarity: 98.636 Percent Identity: 98.636

alignment_block:
US-10-041-770-2 x AF217974 ..
Align seg 1/1 to: AF217974 from: 1 to: 2070
638 ThrProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaPr 654
||||| ||| |||:||||: |||: |||: |||:
3 ACCCCAGGACCCCTCCAGCTCAGGTGCGATCCCCAGATGCCGCCCC 52
654 oProHisProArgThrProLeuGlySerProAlaAlaTyTrpPlyArgV 671
||||| ||| |||:||||: |||: |||: |||:
53 GCCCCATCCAGGACAGCCCTGGGGTCTCCAGCTGCGTACTGGAACAG 102
671 alGlyHisSerAlaCysSerAlaSerCysGlyLysGlyValTrpArgPro 687
||||| ||| |||:||||: |||: |||: |||:
103 TGGGCACTCTGCATGCTCAGCGTCCCTGCGGGAAGGTGCTGGCGCCC 152

```

```

688 lIePheLeuCysIleSerArgGluSerGlyGluLeuAspGluArgSe 704
||||| ||| |||:||||: |||: |||: |||:
153 ATTTTCCTCTGCATCTCCCGTGGTGGAGAGAACTGGATGACGACG 202
704 rCysAlaAlaGlyAlaArgProProAlaSerProGluProCysHisGlyT 721
||||| ||| |||:||||: |||: |||: |||:
203 CTGTGCGGGGTGCCAGGCCGCCAGCTCCCTGAACCTGCAC.GGCA 251
721 hrProCysProProTyTrpGluAlaGlyGluTrpThrSerCysSerArg 737
||||| ||| |||:||||: |||: |||: |||:
252 CCCCATGCCCCCATACTGGGAGGCTGGAGTGGACATCTCGACGCCG 301
738 SerCysGlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPh 754
||||| ||| |||:||||: |||: |||: |||:
302 TCCGTGCCCCCGGACCCAGCCAGCCAGCTGCGCGGCGGCGAGATT 351
754 eGlyGlyGlySerSerValProProGluArgCysGlyHisLeuProA 771
||||| ||| |||:||||: |||: |||: |||:
352 TGGGGGGGTGGCTCTCGGTGCCCGGAGCGCTGTGGACATCTCCCC 401
771 rGProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGlu 787
||||| ||| |||:||||: |||: |||: |||:
402 GCCCAACATCACCGAGCTTGCCAGCTGCCCTCTGTGGCCATGGAA 451
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804
||||| ||| |||:||||: |||: |||: |||:
452 GTTGCTCTCTCTGGAGGACGTGCTCGGTGCGGTGCGGCGGCGGCGAG 501
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821
||||| ||| |||:||||: |||: |||: |||:
502 AAGCGGCGAGTTGCTGTGTGGAAACAGGTGATGAAGTGAGCGAGC 551
821 lngluCysAlaSerGlyProGlnProProSerArgGluAlaCysAsp 837
||||| ||| |||:||||: |||: |||: |||:
552 AGGAGTGTGCTCAGGCCGCCACAGCCCCCAGCAGAGAGGCGCTGTGAC 601
838 MetGlyProCysThrThrAlaTrpPheHisSerAspTrpSerSerLysVa 854
||||| ||| |||:||||: |||: |||: |||:
602 ATGGGGCCCTGTACTACTGCTGTGTTCCACAGCGACTGGAGCTCCAAGT 651
854 lserProGlu 857
652 CTCAGCCCGAG 661
seq_name: gb_pr:BC008840
seq_documentation_block:
LOCUS BC008840 2703 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, hypothetical protein FLJ13710, clone MGC:14171
IMAGE:4120678, mRNA, complete cds.
ACCESSION BC008840
VERSION BC008840.1 GI:14250741
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2703)
Strausberg,R.
Direct Submission
TITLE Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),

```



810 ysValGlyAsnAsnGlyAspGluValSerGluGlnGluCysAlaSerGly 826  
 1220 GTGTGAGCAACATTGGGATGGTGTGACGATGAGGAATGC...AACATG 1266  
 827 ProProGlnProProSerArgGluAlaCysAspMetGlyProCysThrTh 843  
 1267 AACCTCCGGCCGAATGACATGACATGACACTGCGACATGGGACCTGTGCCA 1316  
 843 rAlaTrpPheHisSerAspTrpSerSerLysValSerProGlu 857  
 1317 GAGCTGTTCTCACCAGTGGAGCGAAAGTGCTCAGCGGAG 1359

seq\_name: gb\_pat:E55265

seq\_documentation\_block:  
 LOCUS E55265 5061 bp DNA linear PAT 31-JAN-2002  
 DEFINITION Novel metalloprotease and gene of the same.  
 ACCESSION E55265  
 VERSION E55265.1 GI:18629778  
 KEYWORDS JP 2001008687-A/1.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE  
 1 (bases 1 to 5061)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 AUTHORS Yamaji,N., Nishimura,K. and Sasamata,M.  
 TITLE Novel metalloprotease and gene of the same  
 JOURNAL Patent: JP 2001008687-A 1 16-JAN-2001;  
 YAMANOUCHI PHARMACEUT CO LTD

COMMENT  
 OS Homo sapiens (human)  
 PN JP 2001008687-A/1  
 PD 16-JAN-2001  
 PF 25-JUN-1999 JP 1999180973  
 PR  
 PI NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA  
 PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC  
 C12N9/64, C12Q1/37,  
 PC C12N15/00, C12N5/00

CC  
 FH Key Location/Qualifiers  
 FT source 1..5061  
 FT /organism='Homo sapiens (human)'.  
 FT Location/Qualifiers

FEATURES  
 source 1..5061  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 878 a 1746 c 1587 g 850 t  
 ORIGIN

alignment\_scores:  
 Quality: 807.50 Length: 959  
 Ratio: 1.815 Gaps: 41  
 Percent Similarity: 46.403 Percent Identity: 28.050

alignment\_block:

US-10-041-770-2 x E55265 ..  
 Align seg 1/1 to: E55265 from: 1 to: 5061  
 15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24  
 :|||||: |||  
 596 GTGCTCCAAGCACTGTGGAGTGCAGTGTACCCAGAGCTGGAGTCTCGA 645  
 24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35  
 :|||||: |||  
 646 CGGAGCGGTGGGAGCAGCGGACAGTGGCGCGCCAGCTGAGGGC 695  
 35 hrProThrGluGluGlyGlnGlyProGluGlyValTrpGlyProTrp... 50  
 :|||||: :|||  
 696 TCTACACCGGCGTCGTCACCAAGAGAAGTGGTGGACACCTGGTAG 745  
 50 .....

746 TAGCTGATGCCAAATAGTGGAGTACCACGACAGCGCAGGTTGAGAGC 795  
 51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61  
 :|||||: |||  
 796 TATGTGTGACCATCATGAACATGGTGGCTGTTTCATGACCCAG 845  
 61 lyValGlyVal.....Gln 65  
 :|||||: :|||  
 846 CATTGGGAACCCCATCCACATPACCATGTGGCTGCTGCTGGAAG 895  
 66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro.... 80  
 :|||||: :|||  
 896 ATGAGGAGGAGGACCTAAAGATCAGCCACCATGACACACACCTGAAG 945  
 81 .....SerLeuProLeuP 85  
 :|||||: :|||  
 946 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCTCC 995  
 85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 :||  
 996 CTGACCATG..... 1006  
 102 ProArgProGlnThrSerProGluThrLeuProLeuTyArgThrGlnSe 118  
 :|||||: |||  
 1007 ACATGTCATCTGCTCACCAGAA.....AGGACCTGTGT 1041  
 118 rArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135  
 :|||||: :|||  
 1042 CGAGCCATGAACCGGCTGTGAGACCTGGGACTGCCATGTGGCGGG 1091  
 135 luGluThrGlnGluLeuAlaAlaArgArgSerArgLeuArgAspPro 151  
 :|||||: :|||  
 1092 CATGTGCCAGCGCCAGCCAGCTGAGCATCAGCAGGAGACA..... 1132  
 152 IleLysProGlyMetPheGlyTyArgValProPheAlaLeuProLe 168  
 1132 ..... 1132  
 168 uHisArgAsnArgHisProArgSerProProArgSerGluLeuSerL 185  
 :|||||: |||  
 1133 ....CGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1178  
 185 euLeuSerSerArgGlyGluGluAlaIleProSerProThrProArgAla 201  
 :|||||: |||  
 1179 TTTTGGCATGACATGACGGAAGCGGCAATGACTGTGAGCCCTGGGA 1228  
 202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro.. 215  
 :|||||: :|||  
 1229 AACGACCTTCATCATGCTCTCCACAGCTCTGTACGACGCGCTCCCTC 1278  
 216 .ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232  
 :|||||: :|||  
 1279 ACCTGTCCCGCTGCGCGCCAGTATATACCCAGGTTCCTTGACCGTGG 1328  
 232 exProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248  
 :|||||: :|||  
 1329 GTGGGGCTGTGCTGGACGACCTCTCCCAAGGACATTATCGACTTCC 1378  
 249 LeuArgHisHis.....ProArgAlaGlnAlaSe 258  
 :|||||: |||  
 1379 COTCGTCCACCTGCTGCTCTATGATGTAAGCCACCATGCGGCTC 1428  
 258 rGlyThrGluProProSerPro.....T 266  
 :|||||: |||  
 1429 CAGTACGGGGCTTACTCTGCCTCTCTGCGGAGCATGGATAATGTGCCA 1478  
 266 hrHisSerLeuGlyGlyGlyPhePheArgAlaSerProGlnProArg 282  
 :|||||: :|||  
 1479 CACACTGTGCTGTGGGA.....CCACCTGTC 1510  
 283 ArgProSer....SerGlnGlyTrpAlaSerProGlnValAlaGlyArg... 297  
 :|||||: |||

1511 ACTCCAGCTGGATGACGCCGTGGACGCCACCCGGTGTG...GGGAGAAAT 1557  
298 .....ArgProAspProPheProSerValProArgG 308  
1558 AAGTGTGTCTAGTGGGAGTGGTACCGTGGGCTTCCGGCCCGAGG. 1606  
308 lyArgGlyGlnGlnGlyProTrp.....GlyThrGlyGly 321  
1607 .....CCGTGGATGGTGGCTGGTGGCTGGA 1633  
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338  
1634 GGCCT...GGTCCA..... 1645  
338 pLeuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355  
1646 .....TCTGCTCAGCGAGCTGGCATGGCGGTACA 1676  
355 heAlaProSer...Ser.....ProIleProArg..... 363  
1677 GAGCGCGGAGCGGAGTGCACGAGCGCTACGCCCAATACAAAGGCAGAT 1726  
364 ..CysSerGlyGlySerGluGlnLeuAlaCysSerGlnAlaProCys 379  
1727 ACTGTGGGTGAGCGCAAGCGCTTCGGCTCTGCAACCTGCAGGCGCTGC 1776  
380 ProProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSe 396  
1777 CCGTGGCCACCCCTCTCCGCCACCGTCCAGTGCAGCCACCTTTCACGC 1826  
396 rGlnGluPheMetGlyGlnLeuTrpGlnTyrGlnTrpGluProPheThrGluValG 413  
1827 TATGCTCTACAAGGCCAGCTGCACATCGGTGCCCGTG.....GTCA 1870  
413 lnGlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429  
1871 ATGACGTGAACCCCTGGAGTGCAGTGCCTGCGCGCCGCGAATGAGTACTTT 1920  
430 TyrValArgHisThrGluLysValGlnAsnGlyThrLeuCysGlnPro... 445  
1921 GCCAGAGAGTCGGGACCGCGGTGTCATGGCACCCCTGCTACACAGT 1970  
446 ....GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyC 461  
1971 CCGAGCCAGCGGACCTCTGTCATCAACGCGATCTGTAAAGAGTGGCT 2020  
461 ysAspGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477  
2021 GTGACTTCGAGATTGACTCCGGTCTATGAGGACCCGCTGTGGTGTGTGC 2070  
478 GlyGlyAspSerThrCysArgLeuValSerGlyAsnLeuThrAspAr 494  
2071 CACGGCAACGGCTCCACCTGCCACACCGTGAGCGGACCTTCGAGGAGGC 2120  
494 gGlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaL 511  
2121 CGAGGCG...CTGGGGTATGATGGATGGGGCTGATCCACGCGCGCAC 2167  
511 euArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTrpLeuAlaLeu 527  
2168 GCGAGATCCGATCAAGAGGTGCGGAGGTGCGCACTTCCTGGCACTG 2217  
528 ArgGlyProGly...GlyArgSerIleIleAsnGlyAsnTrpAlaValAs 543  
2218 CGGAGTGAGGACCGGAGAGTACTTCCTCAATGGTGGCTGGACCATCCA 2267  
543 pProProGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnArgP 560  
2268 GTGGAACGGGAGTACCAAGGTGGAGGAGGACCCCTTCATACATCCACGC. 2316  
560 roProArgGluGlyLysGlyGluSerLeuSerAlaGluGlyProThr 576  
2317 .....AGGGCACTGGGAGAACCTCACGTCCCGGGTCCACCC 2355

577 ThrGlnProValAspValTyrMetIlePheGlnGluAsnProGlyVa 593  
2356 AAGGAGCCTGTGTGATCCAGCTGTTCAGGAGAGCAACCTGGGGT 2405  
593 lPheTyrGlnTyrValIleSerSerProProIleLeuGluAsnProT 610  
2406 GCATACAGGTAC..... 2418  
610 hrProGluProProValProGlnLeuGlnProGluIleLeuArgValGlu 626  
2418 ..... 2418  
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGl 643  
2419 .....ACCATCCA 2426  
643 nArglnValArgIleProGlnMetProAlaProProHisProArgThrP 660  
2427 CAGGGAG.....GCAGGTGGCCACGACGAGGTCC 2455  
660 roLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCys 676  
2456 CG.....CCGCCGTGTCTCTCGCATATGGCCCTGGACCAAGTGC 2499  
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693  
2500 ACAGTCACCTGGCGAGAGTGTGCAGAGGCAATGTGTACTGTGTGA 2549  
693 rArgGluSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
2550 GCGGACGAGGAGG...CCGTGGAGGAGGACACTGTGACCCCTGGGCC 2596  
710 rgProProAlaSerProGluProCysHisGlyThrProCysProProTyr 726  
2597 GGCCTGATGACCAACAGAGGAAGTGCAGGACGACGCCCTGCCCTGCCAGG 2646  
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...Gl 742  
2647 TGTGGGACAGTGTGAGTGGCAGCTGTCTCCAGCTCTCGGGCCCTGGGG 2696  
742 TyrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGlyG 758  
2697 CTTCTCCCGCGGCGGTCTGTGATCCGAGCGTGGGCTGGATGAGC 2746  
758 lySerSerValProProGluArgCysGlyHisLeuProArgProAsnIle 774  
2747 AGAGCCCTGGAGCCACCCGCTGTGAACACTTCCCGGCCCTACT 2796  
775 ThrGlnSerCysGlnLeuArgLeu...Cys....GlyHisTrpGluValGl 789  
2797 GAAACCCCTGCAACCGCATGTACCTGTCCGCGCACCTGGGCTGGG 2846  
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806  
2847 GAAC...TGGTCTCAGTGTCACTGACATGTGGGAGGACACTCAGCGCC 2893  
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822  
2894 GAAATGTCTCTGCACCAATGACACCGGTGCCCTGTGACGAGGCCAG 2943  
823 CysAlaSerGlyProProGlnProProSerArgGluAlaCysAspMetGl 839  
2944 .....CAGCCAGCCAGCAAGTCACTGCTCTCTGCC 2975  
839 yProCysThr...ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854  
2976 ACTGTGTGGTGGCCCTGGGCACAC.....TGGGCCCTGAAGGCTCAG 3019  
855 SerProGluProProAlaIleSer 862  
3020 GCAGCGGCTCTCTCAGCCAGCAGC 3043

```

seq_name: gb_pat:AX128433
seq_documentation_block:
LOCUS AX128433 5353 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 2 from Patent WO0131034.
ACCESSION AX128433
VERSION AX128433.1 GI:14134924
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 5353)
AUTHORS Kapeller-Libermann, R. and White, D.
TITLE Protein 27875, a human adam-ts homolog
JOURNAL Patent: WO 0131034-A 2 03-MAY-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
Location/Qualifiers
1..5353
/organism="Homo sapiens"
/db_xref="taxon:9606"
36..5096
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC38921.1"
/db_xref="GI:14134925"
/translation="MPGSPSPAPLRLPRLLLLCALAPGAPGAPGATEGRAALD
IVHVRVDAAGSFUSYELWPRALRRKRVSDAPAFYELQYRGRELRNFNTAHOHL
APGVSTERRRGLGRAHRAHTPACHLLGEVQDPELEGLAAISACDGLGVOLSN
EDYFTEPLDSPAPRQAHPVVKRQAPERLAQKSSAPSTCGVQVYPLESRRR
WEORQWRRLRLRHQSVSKKQVETLVADAKMVEYHGQPOVESVVLTKMNVAG
LHFDPSIGNPHTITRVLVLEDEEDLKITHADNTLKSFKWQKSNMKGDAHPLH
HPTALLPKDLCAAMNRPCTELGISHVAGCQPHRSCSINEDGLPLAFTVAHELGH
SGFHQDSGNDCEPVGKRPIMSPQLLYDAAPLTWRSRQYITRELDKRWGLDLD
PAKDIIDFPPSPGVLDVSHQRLQIGAYSAPCEDMDNVCHTLWCVSVTGCKSLD
AAVDTRGCKNKLSCGCVPGVFPEAPVGGSGWSAWSICSRSGMGVQSAERQCT
QPTPKYKRCYGERKRLCLNQLQAPAGXPSFRVQCSEFDAMLYKQLTWPPVWN
DYNPELHCRPANEFASKLDAVVDVPCYQVRAASRLDINGICKNYGDFEIDSGA
MEDQKCHGNSSTCTHYGTFEEAGLGYVDVGLIPAGAREIQEVAEAAFLALR
SDRKPYLEGWSMTIOWNGDOVAGTETVYARRGNWENTLSPGTPKEPVWQLLFOES
NGCVHYETIHREAGHDEVPVPSWHYGPWKCTVTCGRGVQVQNVYCLERQAGPV
DEHCDPDRDDQKQSEQPCPARWAGEMQLCSCSGGGSLRRVAVLCIRSVGLD
EQSALPEACEHLPPPTPCNRHVPCPATWAGNWSQCSCTGEGTQRNVLCTND
TPVPCDEAQKPASEYTCGLPRLWPLTGLGEGSGSSSHELFEADFIPIHLAPRP
SPASPKQCTMGNAIEEAPLDELPGVVDVDFYDYNFINFHEDLSYGSSEEDLDL
ACTGRTPPHRRPAAPGSPVPATPEAAKEGVLPSPWPSPWSPQASRPPPPSE
CTPGNPLNPLEEDTPIGAPDLPLSLSWPRVSTDLQTPATPESQNDFFVGKDSQS
QLPPWRDRTNVEFDDPEKRGAPHLPPRPSSTLPLSPVSGTHSSPSDVAELWT
GGTVAWEPALGGLGVDSLELMTVGVASLLPPLPAPLPEMKVRSLSLEPTSPFAP
GFGSWDLQTVAYVMTFLPTTLTGLHMEPEALNPGKGPESLTPLEVLSSRLSLTAP
WDSANSHRVPQTPPLASLAPGPPADFLVVRNASQAGNWSCTTCGLGAVWRPY
RCSGGRDECAPAGRPQARRCHLRPCATWHSNWSKGRSCDGGSSVRYQCVQVTRD
LRPLRPHQCPGAPKPAHRCQAPCLSWYTSWREGEACGGGEQORLVTCRPGIL
CEALRPNTTRPCNTPCQVQVWVPGQCSAPCGGVQVORLVKCVNTOTGLPEEDSDQ
CSHEAWPSSRCGTEDECPVEPPRCERDRLSFGFCETLRLGRCQLPTIRTQCCRS
SPSPSGAPSGRQVARR"
BASE COUNT 960 a 1833 c 1648 g 911 t 1 others
ORIGIN
alignment_scores:
Quality: 806.50 Length: 959
Ratio: 1.816 Gaps: 41
Percent Similarity: 46.298 Percent Identity: 28.050
alignment_block:
US-10-041-770-2 x AX128433
Align seg 1/1 to: AX128433 from: 1 to: 5353
15 LeuLeuSerLeuProGlnLeuCys.....LeuAs 24
:::||||: |||
631 GTGCTCCAAGCACCTGTGGAGTGCAAGTGATCCACGAGCTGGAGTCTCGA 680

```

```

24 pGlnGluValLeuSer.....GlyHisSerLeuGlnT 35
|:::|||||
681 CGGGAGCGTGGGAGCAGCGGAGCAGTGGCGGCGCACGCTGAGCG 730
35 hrProThrGluGluGlnGlyProGluGlyValTrpGlyProTirp... 50
:::||||: |||
731 TCTACACGAGCGGTCTGTCACCAAGAGAGTGGGTGGAGACCCCTGGTAG 780
50 .....
781 TAGCTGATGCCAAATATGTTGGAGTACACGAGCAGCGGAGTGGAGAGC 830
51 .....ValGlnTrpAlaSerCysSerGlnProCysG 61
831 TATGTGCTGACCATCATGAACATGTGCTGGCTGCTTTCATGACCCCGAG 880
61 LyValGlyVal.....Gln 65
:::||||:
881 CATTTGGGAACCCATCCACATCATCATTTGCGGCTGCTGCTGCTGGAAG 930
66 ArgArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro..... 80
931 ATGAGGAGGAGGACCTAAAGATCAGCACCATGACGACACACACCCCTGAAG 980
81 .....SerLeuProLeuP 85
981 AGCTTCTGCAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCCATCC 1030
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
||
1031 CCTGCACCATG.....
102 ProArgProGlnThrSerProGlnThrLeuProLeuTyArgThrGlnse 118
|||
1042 ACATCTCCATCTGCTCACCAAG.....AGACCTGTGT 1076
118 rArgArgArgGlyProLeuArgGlyProAlaSerHisLeuGlyArgG 135
::: |||
1077 GCAGCCATCAACGCCCTGTGAGACCCCTGCTCCATCTGCGCGG 1126
135 luGluThrGlnGluLeuArgAlaAlaArgArgSerArgLeuArgAspPro 151
::: |||
1127 CATGTCCAGCGCGCAGCCGAGCTGACAGCTACAGGAGCA..... 1167
152 IleLysProGlyMetPheGlyTyArgValProPheAlaLeuProLe 168
1167 .....
168 uHisArgAsnArgArgHisProArgSerProProArgSerGluLeuSerL 185
|||
1168 ...CGGGCTGCGGCTGCGCTTCACTGTAGCCCGACGAGCTCGGCGACAG 1213
185 eulleSerArgGlyGluGluAlaileProSerProThrProArgAla 201
::: |||
1214 TTTTGGCAITCAGCATGACGGAAGCGCAATGACTGTGAGCCCGTTGGGA 1263
202 GluProPheSerAla.....AsnGlySerProGlnThrGluLeuPro... 215
::: |||
1264 AACGACCTTTCATCATGTCTCCACAGCTCTGTACGACCGCGCTCCCTTC 1313
216 ProThrGluLeuSerValHisThrProSerProGlnAlaGluProLeuS 232
|||
1314 ACTGTGCTCCGCTGAGCGCGCAGTATATACAGGCTTCTTTCACCGCTGG 1363
232 erProGluThrAlaGlnThrGluValAlaProArgThrArgProAlaPro 248
::: |||
1364 GTGGGGCTGTGCTGACGACGACCCCTCTCCAGAGGACATATATGACTTCC 1413
249 LeuArgHis.....ProArgAlaGlnAlaSe 258
|||
1414 CCTCGGTGGCACCTGGCGTCTCTATGATGTAAGCCACGAGTCCCGCTC 1463

```

258 rGlyThrGluProSerPro.....T 266  
1464 CAGTACGGGCTTACTCTCTCTCGAGGACATGATAATGTCTGCCA 1513  
266 hrHisSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArg 282  
1514 CACACTCTGTGTCTGTGGGA.....CCACCTGTCTC 1545  
283 ArgProSer...SerGlnGlyTrpAlaSerProGlnValAlaGlyArg.. 297  
1546 ACTCCAGCTGGATGCGCGGTGAGCGGACCGCGGTG...GGGAGAT 1592  
298 .....ArgProSerProPheProSerValProArg 308  
1593 AAGTGTGTCTCAGTGGGAGTCGACCGGTGGCTTCGGGCCGAGG.. 1641  
308 lyArgGlyGlnGlyGlnGlyProTrp.....GlyThrGlyGly 321  
1542 .....CCGTGGATGTGTGTGTGGTGA 1668  
322 ThrProHisGlyProArgLeuGluProAspProGlnHisProGlyAlaTr 338  
1669 CGGCT...GGTCCA..... 1680  
338 pLeuProLeuSerAsnGlyProHisAlaSerSerLeuTrpSerLeuP 355  
1681 .....TCTGCTACGGAGCTGTGGCATGGCGTACA 1711  
355 heAlaProSer..Ser.....ProIleProArg..... 363  
1712 GAGCGCGAGCGGAGTCACCGAGCTAGCGCCCAATACAAAGCGAGAT 1761  
364 ..CysSerGlyGluSerGlnLeuArgAlaCysSerGlnAlaProCys 379  
1762 ACTGTGTGGTGACGAGCGGCTTCGGCTCTGCAACCTGCAGGCGTGC 1811  
380 ProProGluGlnProAspProAlaLeuGlnCysAlaAlaPheAsnSe 396  
1812 CTGTGTGCGCCCTCTCTCCGCCAGCTCCAGTGCAGCCACTTTCAGC 1861  
396 rGlnGluPheMetGlyGlnLeuTrpGluProPheThrGluValG 413  
1862 TATGCTCTACAGGCGGAGCTGACATGGGTCCCGTG.....GTCA 1905  
413 lnglySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhe 429  
1906 ATGACGTGAACCCCTCGAGTGCACCTCCGCGCCCGGGAATGACTACTT 1955  
430 TyrValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnPro.. 445  
1956 GCGGAGAGCTGCGGAGCGCTGTGATGGCACCCCTCTACCAAGT 2005  
446 ....GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyC 461  
2006 CCGAGCCAGCGGAGCTCTCATCAAGGCACTCTGAAGAAGCTGGCT 2055  
461 ysAspGlyLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477  
2056 GTGACTTCAGATGACTCCGCTCTATGGAGGACCGCTGTGGTGTGTC 2105  
478 GlyGlyAspSerThrCysArgLeuValSerGlyAsnLeuThrAspAr 494  
2106 CAGGCAACGGCTCCACTGCCACACCTGAGCGGACCTTCAGAGGAGC 2155  
494 gGlyGlyProLeuGlyThrGlnLysIleLeuTrpIleProAlaGlyAlaL 511  
2156 CGAGGCG...CTGGGTATGTGGATGTGGGTGATCCACCGCGCGCAC 2202  
511 euArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTrpLeuAlaLeu 527  
2203 GCGAGATCGCATCAAGAGGTTCGCGAGGTGCGCACTTCTGCGACTG 2252  
528 ArgGlyProGly...GlyArgSerIleAsnGlyAsnTrpAlaValAs 543

2253 CGGAGTGGAGACCGGAGAGTACTTCTCTCAATGGTGGCTGGACCATCCA 2302  
543 pProGlySerTyArgAlaGlyGlyThrValPheArgTyArgAsnArgP 560  
2303 GTGGAAACGGGACTACCAAGTGGAGGACCACTTCACATACGACGCG.. 2351  
560 roProArgGluGluGlyLysGlyGluSerLeuSerAlaGluGlyProThr 576  
2352 .....AGGGCACTGGGAGAACCTCACGTCCCGGGTCCCACC 2390  
577 ThrGlnProValAspValTyMetIlePheGlnGluGluAsnProGlyVa 593  
2391 AAGGACCTGTCTGGATCCAGCTGTGTCCAGGAGAGCAACCCCTGGGT 2440  
593 lPheTyGlnTyValIleSerSerProProIleLeuGluAsnProT 610  
2441 GCACACGAGTAC..... 2453  
610 hrProGluProProValProGlnLeuGlnProGluIleLeuArgValGlu 626  
2453 ..... 2453  
627 ProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2454 .....ACCATCCA 2461  
643 nArgGlnValArgIleProGlnMetProAlaProHisProArgThrP 660  
2462 CAGGAG.....GCAGTGGCCACGACGAGGTCC 2490  
660 roLeuGlySerProAlaAlaTyTrpLysArgValGlyHisSerAlaCys 576  
2491 CG.....CCGCCGTGTCTCTGGCATATGGGCCCTGGACCAAGTGC 2534  
677 SerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSe 693  
2535 ACAGTCCCTCGGAGAGTGTGCAGAGGAGCAATGTGTACTGTCTGGA 2584  
693 rArgGluSerGlyGluLeuAspGluArgSerCysAlaAlaGlyAlaA 710  
2585 GCGCAGGAGGAG...CCGCTGACGAGGAGCACTGTGACCCCTGGGCC 2631  
710 rProProAlaSerProGluProCysHisGlyThrProCysProProTy 726  
2632 GGCTATGACCAACAGAGAGTGCAGGAGCGCTGCCCTGCCAGG 2681  
727 TrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGlyPro...G 742  
2682 TGCTGGCAGTGAAGTGGCAGCTGTCTCAGCTCTCGCGGCTGGGG 2731  
742 yThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGly...GlyGly 758  
2732 CTTCTCCCGCGCGCTGTCTGTGATCCGAGCTGGGGCTGGATGAGC 2781  
758 lySerSerValProGluArgCysGlyHisLeuProArgProAsnIle 774  
2782 AGAGCCCTGGAGCCACCGCTGTGAACACCTCCCGGCCCTACT 2831  
775 ThrGlnSerCysGlnLeuArgLeu...Cys...GlyHisTrpGluValG 789  
2832 GAAACCCCTGCAACCGCCATGATCCCTGCCGCGCCACCTGGCTGTGG 2881  
789 ySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerA 806  
2882 GAAC...TGCTCTCAGTCTCAGTGAATGTGGGAGGCACTCAGCGC 2928  
806 rgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGlu 822  
2929 GAAATGTCTCTGCAACCAATGACACCGGTGTCCCTGTGACGAGGCCAG 2978  
823 CysAlaSerGlyProGlnProProSerArgGluAlaCysAspMetG 839



2979 .....CAGCCAGCCAGCGAAGTACACTGCTCTCTGCC 3010  
 839 yProCysThr.ThrAlaTrpPheHisSerAspTrpSerSerLysVal... 854  
 3011 ACTCTGTCGGTGGCCCTGGGCACAC.....TGGGCCCTGAGGCTCAG 3054  
 855 SerProGluProProAlaIleSer 862  
 3055 GCAGCGGCTCCTCCAGCCACGAGC 3078

seq\_name: gb\_pat:AX319851

seq\_documentation\_block:  
 LOCUS AX319851 3312 bp DNA linear PAT 14-DEC-2001  
 DEFINITION Sequence 15 from Patent WO0183782.

ACCESSION AX319851  
 VERSION AX319851.1 GI:17901441

KEYWORDS

SOURCE

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (sites) Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and

AUTHORS

Payne,V.

TITLE Novel proteases

JOURNAL Patent: WO 0183782-A 15 08-NOV-2001;

Sugen, Inc. (US)

FEATURES

Location/Qualifiers

1..3312

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 620 a 1106 c 1035 g 551 t

ORIGIN

alignment\_scores:

Quality: 797.50 Length: 954

Ratio: 1.693 Gaps: 45

Percent Similarity: 49.371 Percent Identity: 28.512

alignment\_block:

US-10-041-770-2 x AX319851 ..

Align seg 1/1 to: AX319851 from: 1 to: 3312

4 TrpThrGlyArgProTrpLeuLeuLeuLeuSerLeuProGlu 20

604 TGGAAAGGGCGCCATGG...TGGCTGGGAGCCTTGAAGCCACCGCTGC 650

20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGlnThrProT 37

651 CAGGCCCTGGGAATGAA.....A 670

37 hrGluGluGlyGln.....GlyProGlu.GlyVa 46

671 CAGAGCGTGGCCAGCCAGGCGCTGAAGCGATCGCTCAGCGAGCGCTAC 720

46 lTrpGlyProTrp..... 50

721 GTGGAGACCTGTGGTGGTGTGACAGATGATGTTGGCCTATCACGGGG 770

51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66

771 CCGGGATGTGGAGCAGTATGCTCCTGCCGTCATGAACATTTGTCACAA 820

67 ArgSerArgThr..... 70

821 TTTTCCAGGACTCGAGTCTGGGAGCACCCTTAACATCCTCGTAATCGC 870

71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85

871 CTCATCCTGCTCAGGAGGAGGAGCCCACTCTGGAGATCA..... 910

85 roProA-gProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101  
 911 ..CCACCATCGCGGAAGTCCCTGGACAGCTTCTGTAAAGTGGCAGAAAT 958  
 102 ProArgProGlnThrSerProGluThrLeuProLeuTrpArgThr 116  
 959 CCATCGTGAACACAGCGGCATGGCAATGCCATTCAGAGAACGGGTGTG 1008  
 117 .....GlnSerArgGlyArgGlyGlyP 124  
 1009 GCTAACCATGACACAGCAGTCTCATCAGCGCTATGACATCTGCATCTA 1058  
 124 roLeuArgGlyProAlaSerHisLeuGlyArgGluThrGlnGluLe 140  
 1059 CAAGAACAACCCCTCGCGCACACTAGCCCTGGCCCCGGTGGCGGAATGT 1108  
 141 ArgAlaAlaArgArg.....SerArgLeuArgAspProIleLysProG 155  
 1109 GTGAGCGGAGAGAGCTGCGCGCTCAATGAGGACATTTGGCTGGCCACA 1158  
 155 yMetPheGlyTyrGlyArgValProPheAlaLeuProLeu.HisArgAsn 171  
 1159 G.....CGTTCACCATTTGCCACGAGAT 1181  
 172 ArgArgHisProArgSerProProArgSerGluLeuSerLeuLeuSerSe 188  
 1182 CGGCACACATTCGCATGAACCA.....TGACG 1210  
 188 rArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPheS 205  
 1211 GCCTGGGAACAGCTGTGG.....GGCCCGTGGTCTCAGGACCCAGCC 1251  
 205 eAlaAsnGlySerProGln...ThrGluLeuProProThrGluLeuSer 220  
 1252 AGCTCATGGTGCACATTAACATGAAGACCAACCCATTCGTGTGGTC 1301  
 221 ValHisThrPro.....SerProGlnAlaGluProLeuSerProGluTh 235  
 1302 ATCCTGACGCGGTACTACATCACCGAGCTTTCTAGACTCGGCGCTGGGC 1351  
 235 rAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgHisH 252  
 1352 TCTGCTGAACAACGCGCCCGCCAGACAGAGACTTGT..... 1388  
 252 isProArgAlaGlnAlaSerGlyThrGluPro..... 262  
 1389 .....GTACCCGACAGTGGCACCGGCCAACGCTACGATGCAGATGA 1430  
 263 ProSerProThrHisSerLeuGlyGlu.....G 272  
 1431 GCAATGCCGCTTTCAGCATGGATGCAAAATCGCTCAGTGAATACGGGG 1480  
 272 yGlyPhePheArgAla.....SerProGlnPro..... 281  
 1481 AGGTCTGCAGCGAGTGTGGTGTCTGTGACGAAGACCAACCGTGCATCAC 1530  
 282 ..ArgArgProSerSerGlnTrpAlaSerProGlnValAlaGlyArg 297  
 1531 AACAGCATCCCGCGCGGAGGCGACGCTGTCCAGACGACACCATCGA 1580  
 298 ArgPro.....AspProPheProSerValProAr 307  
 1581 CAAGGGGTGCTGTACAAACGGGTCTGTCCCTTTGGTGGCGGCCAG 1630  
 307 gGlyArgGlyGlnGlnGlyGlnGlyProTrpGlyThrGlyThrProH 324  
 1631 AGGGTGTGGACGG.....AGCCTGGGGCGCGTGGACT 1662  
 324 isGlyProArgLeuGluProAspProGlnHisProGlyAlaTrpLeuPro 340  
 1663 CCATGGGGGCTGACGCGGAC..... 1685

341 LeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSer.....Le 354  
1686 .....CTGTGGCGGCGGTCTCT 1705  
354 upHeAlaProSer.SerProIleProArg..... 363  
1706 CTCTACTCGCTACTCGACAGCCAGCCCAACCATCGGGGCAAGTAC 1755  
364 CysSerGlyLeuSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380  
1756 TGCTGGGTGAGAGAGCGGCGACCCCTCTGCAACACGATGATGCTCC 1805  
380 OPtrogLInProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerG 397  
1806 CCCTGGCTCCAGGACTTCAGAGAGTGCAGTCTTCTGAATTGACAGCA 1855  
397 InCluPheMetGlyGlnLeuTyrlGlnTrpGluProPheThrGluValGln 413  
1856 TCCTTTCCGTGGGAATCTCAAGTGAACAGTAC...CGGGAGGG 1902  
414 GlySerGlnArgCysGluLeuAsnCysArgProArgGlyPheArgPhety 430  
1903 GCGTGAAGCCCTGCTCGCTACGTCCCTAGCGGAAGGCTTCAACTCTA 1952  
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
1953 CACGGAGAGGGCGGACGCGGTGTGACGGACACCCCTGCCGTCCAGACA 2002  
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463  
2003 CGTGGACATTTGGTCAAGTGGGAGTGAAGCAAGCACCGTGGCTGCGACCA 2052  
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAs 480  
2053 GTCTGGCTCCGACCTCGCGGGAGGACAAGTCCGAGTGTGTGGCGGTGA 2102  
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
2103 CGGAGTGTCTGGAGACATCAGAGGCGTCTTCAGCCACGCTCAGCTG 2152  
497 roLeuGlyTyrlGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeu 513  
2153 GGGCGGGTACGAGGATGTGTGTGATTCCCAAGGCTCCGTCCACATC 2202  
514 GlnIleAlaGlnLeuArgProSerSerAsnTyrlLeuAlaLeuArgGlyPr 530  
2203 TTCATCAGGATCGAACCTCTCTCTCAGTCACTTGGCCCTGAAGGAGA 2252  
530 oGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProProGlys 547  
2253 CCAGGAGTCCCTGCTGTGGAGGGGCTGCCGGGACCCCGCCAGCCACCC 2302  
547 eTyrlArgAlaGlyGlyThrValPheArgTyrlAsnArgProProArgGlu 563  
2303 GTCTGCTCTAGCTGGGACCACTTCTCACTGCGACAGGGGCGGACGAC 2352  
564 GluGlyLysGlyGluSerLeuAlaGluGlyProThrThrGlnProva 580  
2353 GTC.....CAGAGCTCGAAGCCCTGGGACCGATTAAATGATCTCT 2393  
580 lAspValTyrlMetIlePheGlnGluGlnAsnProGlyValPheTyrlGlnT 597  
2394 CATGCTCATGTGTGTGGCCCGGACCGAGCTGCTGCTGCTGCTGCTGCT 2443  
597 yrValIleSerSer.ProProIleLeuGluGlnAsnProThrProGluPr 613  
2444 TC.....AATGCCCCATCGCCGCTGATCGTGCCTCC...CCTACTCC 2484  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
2485 TGGCACTATGGCCCTTGACCAAGTGTGCTGGCCCGGAGTGTGCGGCGTAG 2534  
622 .....Ile.LeuArgValG1 626

2535 CCAGGTGCAGCGGTGGAGTGCAGCAACCCAGCTGGACAGCTCCCGGTGCG 2584  
626 upProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2585 CCCCCCACTACTGAGTGCAGTCCACAGCAAGCT.....GCCC 2619  
643 InArgGlnValArgIleProGln.MetProAlaProProHisProArgTh 659  
2620 AAAAGGCAGCGCTGCACACAGGAGCTTGCCTCCAGAC..... 2661  
659 rProLeuGlySerProAlaAlaTyrlTrpLysArgValGlyHisSerAlaC 676  
2662 .....TGGTGTAGGGAACCTGGTCTGCTCT 2686  
676 ysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIle 692  
2687 GCAGCGCGAGCTGGATCGAGCGTGCAGCGCTCGTGTGTGTGCGAC 2736  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706  
2737 CCGCGCTCTCTGCGCGGAGAGAGCGCTGGACGACGACGCGATGCC 2786  
706 aAlaGlyAlaArgProAlaSerProGluProCysHisGlyThrProC 723  
2787 GCAG...CGCGCCACCTGTACTG...GAGCGCTGCCACGCGCCACTT 2830  
723 ysProProTyrlTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
2831 GCGCTCCGAGTGGCGGCGCTCGACTGTGTGTGAGTGCACCCCGAGTGC 2880  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG1 756  
2881 GGGCGGCGCTCCGCGACCGGTGTGTCTTTCAGAGC.....GCAGA 2924  
756 yGlyGlySerSerValProProGluArgCysGlyHisLeuProArgProA 773  
2925 CCACGCGCGACGCTGCCCGCGCGACTGTCTACCCCGCGCGCACAGCCAC 2974  
773 snIleThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
2975 CGGCGCACATGCGTGCATCTTGCCTGCGCGCTGCCCGCGCGCTGGGTG 3024  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
3025 GCTGGCGAG...TGGGTGAGTGTCTGTGCACAGTGCAGCGTGGCGAGCG 3071  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
3072 GAGCGCTCGGTGCGTGCACGACGACGACGCGGCGAG.....GCGTGC 3115  
821 InGluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837  
3116 ACGAGTGCAGGAGGCC...CTGGCGCGCCCGCCAGCAGCAGCGTGTGAG 3162

seq\_name: gb\_pat:E55273

seq\_documentation\_block: 3312 bp DNA linear PAT 31-JAN-2002  
LOCUS E55273  
DEFINITION Novel metalloprotease and gene of the same.  
ACCESSION E55273  
VERSION E55273.1 GI:18629786  
KEYWORDS JP 2001008687-A/9.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 3312)  
Yanai,N., Nishimura,K. and Sasamata,M.  
AUTHORS Novel metalloprotease and gene of the same  
TITLE Patent: JP 2001008687-A 9 16-JAN-2001;  
JOURNAL YAMANOUCHI PHARMACEUT CO LTD  
COMMENT OS Homo sapiens (human)

```

PN JP 2001008687-A/9
PD 16-JAN-2001
PF 25-JUN-1999 JP 1999180973
PI NOBORU YANAJI, KOICHI NISHIMURA, MIHO SASAMATA
PC C12N15/09.C07K16/40.C12N1/15.C12N1/19.C12N1/21.C12N5/10, PC
C12N9/64.C12Q1/37,
PC C12N15/00.C12N5/00
CC
FH Key 1 Location/Qualifiers
FT source 1..3312 /organism='Homo sapiens (human)'.
FT Location/Qualifiers
1..3312
/organism='Homo sapiens'
/organism='Homo sapiens'
BASE COUNT 522 a 1107 c 1032 g 551 t
ORIGIN
alignment_scores:
Quality: 796.50 Length: 954
Ratio: 1.691 Gaps: 45
Percent Similarity: 49.371 Percent Identity: 28.512
alignment_block:
US-10-041-770-2 x E55273 ..
Align seg 1/1 to: E55273 from: 1 to: 3312
4 TrpThrGlyArgProTrpLeuLeuLeuLeuSerLeuProG1 20
|||||
604 TGAAGGCGGCCCATGG...TGCTGCGGACCTTGAAGCCACCGCTGC 650
20 nLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuGlnThrProt 37
|||||
651 CAGGCCCTGGGGAATGA.....A 670
37 hrGluGluGln.....GlyProGlu.GlyVa 46
|||||
671 CAGACGGTGGCCAGCCAGCGCTGAAGCGATCGGTACCGAGAGCGCTAC 720
46 IrpGlyProTrp..... 50
721 GTGAGACCTTGGTGGCTGACAAAGATGATGATGCGCTATCACGGGCG 770
51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66
|||||
771 CCGGATGTGAGCAGATATGCTCTGGCCATCATGAACATTGTGCCAACC 820
67 ArgSerArgThr..... 70
821 TTTTCAGGACTCGAGTCTGGGAAGCACCGTTAAATCCTCGTAACTCGC 870
71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeuP 85
|||||
871 CTCATCTCTGCTCAGGAGGAGCCACCCACTCTGGAGATCA..... 910
85 roProArgProProArgHisProGlnAlaLeuLeuProArgGlyGlnGly 101
|||||
911 ..CCCAACCATCGCGGAAGTCCCTAGACAGCTTCTGTAAGTCGAGAAAT 958
102 ProArgProGlnThrSerProGluThrLeuProLeuTyraThr..... 116
|||||
959 CCATCGTGAACACAGCGGCCCATGCCATGCCATTCACAGAACGGTGTG 1008
117 .....GlnSerArgGlyArgGlyGlyP 124
|||||
1009 GCTAACCATGACACAGCAGTCTCATCACACGCTATGACATCTGCATCTA 1058
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluIle 140
|||||
1059 CAGAACAAACCTTGGCGCACCATAGGCTTGGCCCTGGCCCGCGGGAATGT 1108

```

```

141 ArgAlaAlaArgArg.....SerArgLeuArgAspProIleLeysProG1 155
|||||
1109 GTGACCGCGAGAGAGCTGCAGCTCAATGAGGACATGGCTGGCCCA 1158
155 YmetPheGlyTyrglyArgValProPheAlaLeuProLeu.HisArgAsn 171
|||||
1159 G.....CGTTCACCATTTGCCACAGAT 1181
172 ArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSerSe 188
|||||
1182 CGGCGACACATCGGCATGAACCA.....TGACG 1210
188 rArgGlyGluGluAlaIleProSerProThrProArgAlaAlaProPheS 205
|||||
1211 CGGTGGGAACAGCTGTGG.....GGCCGTGGTGCAGGACCCAGCC 1251
205 erAlaAsnGlySerProGln...ThrGluLeuProProThrGluLeuSer 220
|||||
1252 AGCTCATGGCTGCCACATTACCATGAAGACCAACCCATCGTGTGTC 1301
221 ValHisThrPro.....SerProGlnAlaGluProLeuSerProGluTh 235
|||||
1302 ATCTGTGACCGCTGACTACATCACAGCTTCTAGACTGGGCTGGGGC 1351
235 rAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArgHisH 252
|||||
1352 TCTGCTGACACACCGGCCGCCAGACAGGACTTGT..... 1388
252 IsProArgAlaGlnAlaSerGlyThrGluPro..... 262
1389 .....GTACCGACAGTGGCCGCGGCAAGCCCTACGATGATGA 1430
263 ProSerProThrHisSerLeuGlyGlu.....G1 272
|||||
1431 GCATGCGCGCTTTCAGCATGGAGTCAATCGCTCAGTCAATACGGGG 1480
272 yGlyPhePheArgAla.....SerProGlnPro..... 281
|||||
1481 AGGTCTGACGAGCTGGTGTCTGAGCAAGAACCCGTCGATCAACC 1530
282 ..ArgArgProSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArg 297
|||||
1531 AACAGATCCCGCGCGCGAGGCGACGCTGTGCCAGACGACACCATCGA 1580
298 ArgPro.....AspProPheProSerValProAr 307
|||||
1581 CAGGGTGTGTCTACAAACGGGTCTGTCTCCCTTTGGTGGCGCCAG 1630
307 gGlyArgGlyGlnGlnGlyGlnGlyProTrpGlyThrGlyGlyThrProH 324
|||||
1631 AGGTTGTGGACGG.....AGCCTGGGGCGGTGGACT 1662
324 IsGlyProArgLeuGluProAspProGlnHisProGlyAlaTrpLeuPro 340
|||||
1663 CCATGGGCGGACTGCAGCCGGAC..... 1685
341 LeuLeuSerAsnGlyProHisAlaSerSerLeuTrpSer.....Le 354
|||||
1686 .....CTGTGGCGGCGGTGTGCT 1705
354 uPheAlaProSer.SerProfileProArg..... 363
|||||
1706 CTTTAGCCGTCTACTGGACAGCCCGCCAGCCACCATCGGGGCAAGTAC 1755
364 CysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysPr 380
|||||
1756 TGTCTGGTGAAGAAGGCGGCGACCGCTCTCTGCAACACGATGACTGTCC 1805
380 oProGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerg 397
|||||
1806 CCTGGCTCCCGAGACTTCAGAGAAAGTGCAGTGTTCGAAATTTCACAGCA 1855

```

397 lngluPheMetGlyGlnLeuTyrGlnTrpGluProPheThrGluValGln 413  
1856 TCCTTTTCGGTGGAAATCTCAAGTGGAAACGTAC...CGGGAGGG 1902  
414 GlySerGlnArgCysGlnLeuAsnCysArgProArgGlyPheArgPheTy 430  
1903 GGGTGAAGCCCTGCTCGCTCAGCTAGCGGAAGGCTTCAACTCTA 1952  
430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
1953 CACGAGAGGGCGGCGAGCGTGTGTGACGGGACACCTGCCGTCAGACA 2002  
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGly 463  
2003 CGGTGGACATTTCGTCAGTGGCGAATGAACACCGTGGGTGGCGACCA 2052  
464 IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyLys 480  
2053 GTCTGGGCTCCGACCTGCGGGAGAGCAAGTCCGAGTGTGTGGCGGTGA 2102  
480 pAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
2103 CGGCAGTGCCTGGAGACCATCGAGGGCTCTTCAGCCCGACGCTCACCTG 2152  
497 roLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeu 513  
2153 GGGCGGGGTACGAGGATGCTGCTGGATTCCCAAGGCTCCGTCCACATC 2202  
514 GlnIleAlaGlnLeuArgProSerAsnTyrLeuAlaLeuArgGlyPr 530  
2203 TTCACAGAGCTGACACCTCTCTCAGTCACCTTGGCCCTGAGGGAGA 2252  
530 oGlyGlyArgSerIleLeuAsnGlyAsnTrpAlaValAspProGlyLys 547  
2253 CCAGGAGTCCCTGCTGTGGAGGGCTGCGGGAGACCCCGCCAGCCACCC 2302  
547 eTyrArgAlaGlyThrValPheArgTyrAsnArgProProArgGlu 563  
2303 GTCGCTCTAGCTGGGACCACTTTCACCTGCACAGGGGCGCCAGACCAG 2352  
564 GluGlyLysGlyGlyLeuSerLeuAlaGluGlyProThrThrGlnProVa 580  
2353 GTC.....CAGAGCTCGAAGCCCTGGACCGATTAAATGATCTCT 2393  
580 lAspValTyrMetIlePheGlnGluLysAsnProGlyValPheTyrGlnT 597  
2394 CATGCTCATGCTGTGCGCGGACCGAGCTGCTGCTGCTGCTGCTGCTGCT 2443  
597 yrValIleSerSer.ProProIleLeuGluAsnProThrProGluPr 613  
2444 TC.....AATGCCCATCGCCGCTGACTGCTGCTGCTGCTGCTGCTGCT 2484  
613 o.....ProValProGlnLeuGlnProGlu..... 621  
2485 TGGCACTATGCGCCCTGACCAAGTGTGCGCCCAAGTGTGCGCGGTAG 2534  
622 .....Ile.LeuArgValG 626  
2535 CCAGTGGAGCGGTGGAGTCCCGACACCGAGCTGGACAGCTCCGCGGTG 2584  
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2585 CCCCCCACTACTGAGTGCACACAGCAAGCT.....GCC 2619  
643 lncArgGlnValArgIleProGln.MetProAlaProProHisProArgTh 659  
2620 AAAGGACGCGCGCTGCAACAGGAGGCTTGCCTCCAGAC..... 2661  
659 rProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHisSerAlaC 676  
2662 .....TGGTTGTAGGGAAGTGGTGGCTCT 2686  
676 ysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIle 692

2687 GCAGCGCGAGCTGGATCGAGCGTGGCGAGCGCTGGTGGTGGCCAG 2736  
693 SerArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAl 706  
2737 CGCCGCTCTCTGCGCGGAGAGGAGCGGTGGAGACGACGCGATGCC 2786  
706 aAlaGlyAlaArgProProAlaSerProGluProCysHisGlyThrProC 723  
2787 GCAG...CGCGCGCCACCTGTACTG...GAGCGCTGCCAGCGCCCACTT 2830  
723 ysProProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys 739  
2831 GCGCTCCGAGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2880  
740 GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyG 756  
2881 GGGCGGGGCTCCGCGACCGGTGCTTTCGAAGAGC.....GCAG 2924  
756 yGlyGlySerValProProGluArgCysGlyHisLeuProArgProA 773  
2925 CCACCGCGCAGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2974  
773 snIleThrLysSerCysGlnLeuArgLeuCys.....GlyHisTrpGlu 787  
2975 CGCGCACCATGCTGCACTTTCGCGCTGCGCGCGCGCGCTGGGTG 3024  
788 ValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnAr 804  
3025 GCTGGCGAG...TGGGGTGGTGGTCTGCACAGTGGCGCTGGCGCAGCG 3071  
804 gSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluG 821  
3072 GAGCGCTGGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3115  
821 lngluCysAlaSerGlyProProGlnProProSerArgGluAlaCysAsp 837  
3115 ACAGTGGCAGGAGGCG...CTGCGCGCGCGCGCGCGCGCGCGCGCGCG 3162

seq\_name: gb\_pr:AF163762

seq\_documentation\_block:  
LOCUS AF163762 3400 bp mRNA linear PRI 01-DEC-2000

DEFINITION Homo sapiens zinc metalloendopeptidase (ADAMTS10) mRNA, partial cds.

ACCESSION AF163762  
VERSION AF163762.1 GI:11493588

KEYWORDS SOURCE  
human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
ADAM-TS10: A novel member of the ADAM-TS family containing multiple thrombospondin type I repeats

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
Apte,S.S.  
Submitted (29-JUN-1999) Biomedical Engineering, Lerner Research  
Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue,  
Cleveland, OH 44195, USA

FEATURES  
Location/Qualifiers  
1..3400

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..3400  
/gene="ADAMTS10"

gene  
/note="a disintegrin-like and metalloprotease domain with  
thrombospondin type I repeats-10"  
<1..3234  
/gene="ADAMTS10"

CDS  
/gene="ADAMTS10"

```

/codon_start=1
/product="zinc metalloendopeptidase"
/protein_id="AAG35563.1"
/db_xref="GI:11493589"
/translation="RSODELSLESVEIAFPTRVDHNGALLAFSPPPRRRGRTGA
TAGSRLYKYVASPSTHPLNLNTRSRLLAGHSVVEYTRBGLAQWRAARPHCLYAGHL
QGQASSHVAISCGGLHGLIVADEEYLIEPLHGPKGSRPESGPHVYKRSRL
LHDHDTACGVDERKPKRWPLNLTLPAPPLGNTERGQGLKARSVRERIVETL
VADKMMVAYHGRDVEQYVLAIMNIVAKLFQDSLSGSTVNIILYLLILTEDQPLE
ITHHAGKSLDFCKWKQKIVNHSHGNAIPENGVAHNDTAVLTIRYDICTYKKNPCGT
LGLAPVGMGERSCSNVEDIGLPQAFTHAEGHTFGMHDGVGNSCGARQDPFAK
LMAHITMKNTPFVMSKNEDYITSPLDGLGLCLNNRPPQDFVPTVAPQAYDAD
EQCFQHGKSRQCKYGEVSELWCLSKSNRCITNSTPAEGLTCOTHTDKWCYKR
VCFPGSRPEVGDWAGPFWGDCSRTCCGGVSSSRHCDSPRPIGGKYCLGERRR
HRSNCTDDCPGSDREVQCSSEFDSIPFGKFKWIKYRGSGVKACSLSLAEGNF
YTERAAAVDGTPCRPDPTDVCYSGECHKVCDRLGSDLDREDKRCVCGDGSACETI
EGVFPSPAGAYEDVYVPIKGSYHIFIQDLNLSHLALKQDSLELLLEGLPFPQ
HRLPLAGTTFOLRGQPDQVOSLEALPINASLIYVNLARTLPAIRYFNAPLARDSL
PPYSWHVAPWTKCSAQACGGSVOVAECRNQLOSSAVAPHYCSAHSKLPKRQACNTE
PCPDVWVGNLSLCSRCODAGVRSVVCORRYSAAREKALDDSDACPPRPPVLEACH
GPTPPPEWALDNSECTPSCGPGLRHVRVICKSADHRATLPPAHCSAAKPPATMRCN
LRRCPFAHWVAGEWGECSAQCGVGGRVSRVCTSHTGQASHECTEALRPPTTQOCEAK
CDSPTGDPGPECKDNKVAICPLVLKFCQFSRAFYFROMCKCTCGH"
BASE COUNT      646 a 1117 c 1083 g 554 t
ORIGIN

```

```

alignment_scores:
  Quality: 791.50      Length: 953
  Ratio: 1.666         Gaps: 43
  Percent Similarity: 49.843  Percent Identity: 28.122

```

alignment block:

```

US-10-041-770-2 x AF163762
Align seg 1/1 to: AF163762 from: 1 to: 3400

4 TrpThrGlyArgProTrpLeuTyLeuLeuLeuLeuSerLeuProgl 20
526 TGGAAAGGGCGCCATGG...TGGCTCGGACCTTGAAGCCACCGCTGC 572
20 nLeuCysLeuAspGlnGluValLeuSerGlyHisLeuGlnThrProt 37
573 CAGACCCCTGGGAATGAA...A 592
37 hrGluGluGlyGln...GlyProGlu.GlyVa 46
593 CAGAGCGTGGCCACCCAGCGCTGAAGCGATCGGTACAGCCGAGAGCGTAC 642
46 lTrpGlyProTrp... 50
643 GTGGAGACCCCTGGTGGTGTGACAGATGATGGTGCCTATCACGGGCG 692
51 ..ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArg 66
693 CCGGATGTGGAGCAGTATGCTTGGCCATCATCAACATTGTTCCCAAAC 742
67 ArgSerArgThr... 70
743 TTTTCCAGGACTCGAGTCTGGGAAGCACCCTTAACATCCTCGTAACTCGC 792
71 .....CysGlnLeuProThrValGlnLeuHisProSerLeuProLeup 85
793 CTCATCTCTGCACGGAGGACACGCCCACTCTGGAGATCA..... 832
85 roProArgProProArgHisProGluAlaLeuLeuProArgGlyGlnGly 101
833 ..CCCAACCATCGCGGGAAGTCCCTAGACAGCTTCTGTAAAGTGGCAGAAAT 880
102 ProArgProGlnThrSerProGlnThrLeuProLeuTyArgThr..... 116
881 COATCGTGAACACACAGCGGCCATGGCAATGCCATCCAGAACACGGTGTG 930

```

```

117 .....GlnSerArgGlyArgGlyGlyP 124
931 GCTAACCATGACACAGCAGTGTCTCATCAGCGTATGACATCTGCATCTA 980
124 roLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluLe 140
981 CAAGAACAACCCCTGGCGCACACTAGCTGGCCCGCGTGGCGGGAATGT 1030
141 ArgAlaAlaArgSerArgLeuArgAspProIleLeuProGlyMetPh 157
1031 GTGACGGCGGAGAAGCT..... 1048
157 eGlyTyArgValProPheAlaLeuProLeuHisArgAsnArgArgH 174
1049 .....GCAGCGCTCAATGAGGACATTTGGCCTGCC 1076
174 isProArgSerProProArgSerGluLeuSerLeuLeuSerSer..... 188
1077 ACAACGGTTCACCATTTGCCACGAGATGGCGCACACATTCGGCATGAAC 1126
189 ...ArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPh 204
1127 ATGACGGCGTGGAAACACACTGTGGGGCCCGTGGTCAGGACCCAGCCA... 1174
204 eSerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerV 221
1175 ...AGCTCATGGCTGCCACATTAACCATGAAGACCAACCATTCGTGTG 1220
221 alHisThrProSerProGlnAlaGluProLeuSerProGluThrAlaGln 237
1221 GTCAT.....CCTGCAACCGTGACT 1240
238 ThrGluValAlaProArgThrArg.ProAlaProLeuArgHisHisProA 254
1241 ACATCACCAGCTTCTAGACTCGGCGCTGGGCTCTGCTGAACAACCGG 1290
254 rgAlaGlnAla.....SerGlyThrGluPro..... 262
1291 CCCCCAGACAGACTTGTGTACCCGACAGTGGCACCGGGCCAAAGCCTA 1340
263 .....ProSerProThrHisSerLeuGlyGlu..... 271
1341 CGATGCAGATGAGCAATGCCGCTTTCAGCATGGAGTCAAAATCGCTCAGT 1390
272 .....GlyGlyPheArgAla.....SerProGlnP 281
1391 GTAATACGGGAGGTCTCGACGAGCTGTGGTGTCTGAGCAGAGACAC 1440
281 ro.....ArgArgProSerSerGlnGlyTrpAlaSerProGln 293
1441 CGGTGCATCACCAACAGCATCCGCGCCGCGAGGCGACGCTGTGCAGAC 1490
294 ValAlaGlyArgArgPro.....AspProPhePr 303
1491 GCACACCATCGACAGGGGTGTGTACAAACGGGTCTGTGTCCCTTGG 1540
303 oSerValProArgGlyArgGlyGlnGlnGlyProTrpGlyThrG 320
1541 GFTCGCGCCACAGAGGTGTGGACGG.....AGCCTGG 1572
320 lyGlyThrProHisGlyProArgLeuGluProAspProGlnHisProGly 336
1573 GGCCCTGGACTCCATGGGCGGACTCGACGGGAC..... 1607
337 AlaTrpLeu.ProLeuLeuSerAsnGlyProHisAlaSerSerLeuTrps 353
1608 CTGTGGCGGGCGGTCTCTCTCTAGTGTCTACTGC..... 1644
353 erLeuPheAlaProSerSerProIleProArg.....Cys 364
1645 .....GACAGCCCGAGGCCAACCATCGGGGGCAAGTACTGT 1680
365 SerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProPr 381

```

```

||||| ..... |||||||
1681 CTGGGTGAGAGAGCGGCACCGCTCTGCAACAGGATGATGCTGCCCC 1730
|||||
381 oGluInProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnG 398
|||||
1731 TGCGTCCAGGACTTCAGAGAGTGCAGTGTCTGAATTTGACAGATCC 1780
|||||
398 luPheMetGlyGlnLeuTyrlGlnTrpGluProPheThrGluValGlnGly 414
|||||
1781 CTTTCGCGTGGAAATTTCTCAAGTGGAAACGATAC...CGGGAGGGGCG 1827
|||||
415 SerGlnArgCysGluLeuAsnCysAlaArgGlyPheArgPheTyrlVa 431
|||||
1828 GTAAGAGCGCTGCTGCTACGAGCCTAGCGAAGGCTTCAACTTCTACAC 1877
|||||
431 larGHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyAlaP 448
|||||
1878 GGAGAGGGCGGACCGCTGTGTGAGGGGACACCTGCCGTCACAGACGG 1927
|||||
448 roAspIleCysValAlaGlyArgCysLeuSerProGlyCysAspGlyIle 464
|||||
1928 TGGACATTTGCGTCAGTGGCGAATGCAAGCACGTCGGCTGCGACCGAGTC 1977
|||||
465 LeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyAspAs 481
|||||
1978 CTGGCTCCGACCTGCGGGAGGACAGTGCAGTGTGTGGCGGTGACGG 2027
|||||
481 pSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProL 498
|||||
2028 CAGTGCCTGGAGACCATTCAGGGCGCTTCAGCCACCGCTCACCCTGGGG 2077
|||||
498 euGlyTyrlGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGln 514
|||||
2078 CCGGGTACGAGGATGCTGTGGATTCCCAAGGCTCCGTCCACATCTTC 2127
|||||
515 IleAlaGlnLeuArgProSerSerAsnTyrlLeuAlaLeuArgGlyProG 531
|||||
2128 ATCCAGGATCGAACCTCTCTCAGTCACATGTGCCCTGAAGGAGACCA 2177
|||||
531 yGlyArgSerIleIleAsnGlyAsnTrpAlaValAspProGlySert 548
|||||
2178 GGAGTCCTGTGTGGAGGGGTGCTGGGACCCCGCCAGCCCGTC 2227
|||||
548 yrArgAlaGlyGlyThrValPheArgTyrlAsnArgProProArgGluGlu 564
|||||
2228 TGCCTTAGCTGGGACCACTTTCAATCGACAGGGGCCAGACCGATC 2277
|||||
565 GlyLysGlyGluSerLeuSerAlaGluGlyProThrThrGlnProValas 581
|||||
2278 .....CAGAGCTCGAAGCCCTGGGACCGATTAAATGATCTCTCAT 2318
|||||
581 pValTyrlMetIlePheGlnGluGluAsnProGlyValPheTyrlGlnTy 598
|||||
2319 CGTCAATGTGTGGCGGACCGAGCTGCTGGCCCTCGCTACCGCTC. 2367
|||||
598 alIleSerSer.ProProProIleLeuGluAsnProThrProGluPro.. 613
|||||
2368 .....AATGCCCATCGCCGTGATGCTGCTGCC...CCTACTCCTGG 2409
|||||
614 .....ProValProGlnLeuGlnProGlu..... 621
|||||
2410 CACTATGGCCCTGGACCAAGTGTGCGCCCGAGTGTGAGCGGTAGCCA 2459
|||||
622 .....IleLeuArgValGluPr 627
|||||
2460 GTGCGAGGGTGGAGTCCGCAACCACTGGACAGTCCCGGTGCGCC 2509
|||||
627 oProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnA 644
|||||
2510 CCCTACTGTGAGTCCGACAGCAAGCT.....GCCCAA 2544
|||||
644 rgGlnValArgIleProGln.MetProAlaProProHisProArgThrPr 660
|||

```

```

2545 AGCGAGCGCGCTGCAACACGAGCGCTTGCCTCCAGAC..... 2583
|||||
660 oLeuGlySerProAlaAlaTyrlTrpLysArgValGlyHisSerAlaCys 677
|||||
2584 .....TGGTTGTAGGAACCTGGCTCGCTCGCA 2611
|||||
677 erAlaSerCysGlyLysGlyValTrpArgProIlePheLeuCysIleSer 693
|||||
2612 GCCGAGCTGGATGCGAGCGTGCAGCTGCTGCTGCTGTCGACGCGC 2661
|||||
694 ArgGluSerGlyGluGlu.....LeuAspGluArgSerCysAlaAl 707
|||||
2662 CCGCTCTCTGCGCGGAGAGAGCGCTGACGACAGCAGCGATGCCCGCA 2711
|||||
707 aglyAlaArgProProAlaSerProGluProCysHisGlyThrProCysP 724
|||||
2712 G...CCGCGCCCACTGTACTG...GAGGCTGCCAGCGCCCACTGCC 2755
|||||
724 roProTyrlTrpGluAlaGlyLysGlyThrSerCysSerArgSerCysGly 740
|||||
2756 CTCCGAGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2805
|||||
741 ProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757
|||||
2806 CCGGCGCTCCGCGCACCGCTGCTGCTTTCAGAGAC.....GCAGACCA 2849
|||||
757 yGlySerSerValProGluArgCysGlyHisLeuProArgProAsnI 774
|||||
2850 CCGCGCACGCTGCCCGCGCACTGTCTACCCCGCGCGCAAGCCCGG 2899
|||||
774 leThrGlnSerCysGlnLeuArgLeuCys.....GlyHisTrpGluVal 788
|||||
2900 CCACATGCGCTGCACTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTG 2949
|||||
789 GlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSe 805
|||||
2950 GCGCAG...TGGGTGAGTGTCTGCACAGTGGCGCTGCGCGCAGCGCA 2996
|||||
805 rArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnG 822
|||||
2997 GCGCTCGGTGCGCTGCACCGCACCGGCGCAG.....GCGTGCACG 3040
|||||
822 luCysAlaSerGlyProGlnProProSerArgGluAlaCysAsp 837
|||||
3041 AGTGACGAGGAGGC...CTGCGCGCCCGCACCGAGCAGTGTGTAG 3084
|||||

```

seq\_name: gb\_pr:AK023772

seq\_documentation\_block: 2964 bp mRNA linear PRI 29-SEP-2000  
 LOCUS AK023772  
 DEFINITION Homo sapiens CDNA FLJ13710 fis, clone PLACE2000373, weakly similar to F-SPONDIN PRECURSOR.

ACCESSION AK023772  
 VERSION AK023772.1 GI:10435805  
 KEYWORDS cllgo capping, fis (full insert sequence).  
 SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib.PLACE2 clone:PLACE2000373.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fuji,A., Hara,H., Tanase,T., Nomura,Y., Togiyasu,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.  
 NEDO human cDNA sequencing project

JOURNAL  
 REFERENCE 2 (bases 1 to 2964)

AUTHORS Isogai,T. and Otsuki,T.

TITLE Direct Submission  
 JOURNAL Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan







430 rValArgHisThrGluLysValGlnAspGlyThrLeuCysGlnProGlyA 447  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
1953 CAGGGAGAGGGCGCACCCTGTGGTGGACGGGCACACCTCCGTCCAGACA 2002  
447 laProAspIleCysValAlaGlyArgCysLeuSerProGlyCysaspGly 463  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2003 CGGTGGACATTGGCTCAGTGCGGAATGCAAGACACGTGGGCTCGCACCGA 2052  
464 lleLeuGlySerGlyArgArgProAspGlyCysGlyValCysasGlyAs 480  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2053 GTCTTGGGCTCCGACCTCGCGGAGGACAGTCGCCGAGTGTGTGGCGGTGA 2102  
480 AspSerThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyP 497  
\_ :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2103 CGGAGTGCCTTCGAGACCATTGAGGGCGTCTTCAGCCCAGGCTCACCTG 2152  
497 toLeuGlyTrGlnLysrIleLeuTripleProAlaGlyAlaLeuArgLeu 513  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2153 GGGCGGGTAGCAGGATGCTGCTGTGGATTCCCAAAGGCTCCGTCCACATC 2202  
514 GlutlleAlaGlnLeuArgProSerSerAsnTyrlleAlaLeuArgGlyPr 530  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2203 TTCATCCAGGATCTGAACCTCTCTCAGTCACTTGGCCCTGAAGGGAGA 2252  
530 oGlycylArgserIleleAsnGlyAsnTpAlaValAspProProGlyS 547  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::  
2253 CCAGGAGTCCCTGCTGTGGAGGGGCTGCTGGGACCCCCCAGCCCCACC 2302  
547 erTyArgAlaGlyGlyThrValPheArgTyrrAsnArgProProArgGlu 563  
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 13:01:30 ; Search time 2031.65 Seconds  
(without alignments)  
17498.575 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 agggagaactggactggcag.....cctcagccttcacgcatag 2634

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_gss:\*

13: em\_gss\_hum:\*

14: em\_gss\_inv:\*

15: em\_gss\_pln:\*

16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	652.6	24.8	891	9	AL513944 AL513944
2	497.4	18.9	744	10	BF536373 602051728
3	449	17.0	597	9	AI683106 tx01d11.x
4	419.6	15.9	865	10	BI094238 602860118
5	411	15.6	437	10	BM148104 TCAAP1Q83
6	391.4	14.9	517	10	BE808908 214215 MA
7	373.4	14.2	495	9	AW230357 uc65a11.y
8	370.8	14.1	625	10	BG438243 602490152
9	359.8	13.7	866	10	BI330332 602983024
10	357.4	13.6	360	10	BF993048 ILS-GN017
11	338	12.8	592	10	BM391602 UI-R-DY0-
12	337	12.8	370	10	BG005397 PM1-GN018
13	330.8	12.6	513	9	AW763191 ur60h05.y
14	317.8	12.1	375	10	BG008790 MR3-GN018
15	314.4	11.9	324	10	BF993027 ILS-GN017
16	305.2	11.6	449	10	BG900891 HOA7-1-H1
17	303.8	11.5	373	10	BF987707 QV0-GN014

18	287.2	10.9	307	10	BF989747
19	282	10.7	309	9	AI637480
20	279	10.6	327	10	BF993841
21	268.2	10.2	286	10	D78761 HUM510H03B
22	261.4	9.9	507	10	BF986444 QV4-GN014
23	260.8	9.9	571	10	BE757541 212006 MA
24	260.4	9.9	262	10	BF989357 ILS-GN017
25	239.2	9.1	266	10	BM150365 TCAAP9Q92
26	221.4	8.4	498	9	BB855583 BB855583
27	219	8.3	274	10	BM149824 TCAAP3E92
28	214.4	8.1	273	10	BF997426 PM1-GN018
29	205.6	7.8	711	10	BI836313 603083995
30	195.2	7.4	389	10	BG008820 MR3-GN018
31	193	7.3	133	10	BF870887 ILO-ET015
32	183.2	7.0	284	10	BF997840 MR2-GN012
33	182.2	6.9	216	10	BF997839 MR2-GN012
34	181.6	6.9	380	9	AI521594 to65c12.x
35	176	6.7	214	10	BF997840 MR2-GN012
36	170.8	6.5	183	10	BF997839 MR2-GN012
37	159.8	6.1	775	10	BI645713 603275026
38	153.6	5.8	479	10	BF443732 261436 MA
39	146.4	5.6	480	10	BF651524 274324 MA
40	146.2	5.6	446	9	BB862174 BB862174
41	140	5.3	510	10	BF412563 UI-R-BT1-
42	139.8	5.3	379	9	BB857367 BB857367
43	137	5.2	483	9	BB856356 BB856356
44	134.6	5.1	664	12	AZ335574 LM0065P2Z
45	134	5.1	691	9	BB643815 BB643815

ALIGNMENTS

AL513944 891 bp mRNA linear EST 13-FEB-2001

AL513944 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CLOBA010ZE04 5

prime, mRNA sequence.

AL513944

AL513944.1 GI:12777438

EST.

ORGANISM human.

REFERENCE 1 (bases 1 to 891)

1 (bases 1 to 891)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 891

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CLOBA010ZE04"

/clone\_lib="LTI\_NFL006\_PL2"

/tissue\_type="placenta"

/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(gt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 162 a 305 c 265 g 155 t 4 others

ORIGIN

```

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM9519 row: g column: 08
High quality sequence stop: 685.

FEATURES             Location/Qualifiers
     source            1..744
     organism="Mus musculus"
     strain="FVB/N"
     db_xref="taxon:10090"
     clone="IMAGE:4191007"
     clone_lib="NCI_CGAP_SG2"
     lab_host="DH10B (T1 phage-resistant)"
     note="Organ: salivary gland; Vector: pCMV-SPORT6; Site:1.
NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      139 a      231 c      166 t
ORIGIN
Query Match      18.9%; Score 497.4; DB 10; Length 744;
Best Local Similarity 79.3%; Pred. No. 3.5e-92;
Matches 618; Conservative 0; Mismatches 126; Indels 35; Gaps 1;

QY 1281 cgcctctatgtcgtccaaactgaaaggtccaggatgggacctgtgtcagcctgagc 1340
      |||||
Db 1 CGGATTCTATGTCGACACACTGAAAGGTGCAGGATGGAACCTGTGTCCAGCTGGATC 60

QY 1341 cctcgacatctgtgtggtcgagcgtgtctgagcccgctgtgagtggatccttggctc 1400
      |||||
Db 61 CCTAGACATCTGTGTGGCTGGACGCTCCCT----- 90

QY 1401 tggcaggcgtcctgatggctgtgtgagtcgtgtggggtgatgattacactgtgccttgt 1460
      |||||
Db 91 -----GGCGTCCGGATGGCTTCGCGAGTCTCGGGGGGTGATGGTCTACCTGCCGCGTGGT 145

QY 1461 ttcggggaacctcactgaccagagggggcccccctgggctatcagaagatcttggattcc 1520
      |||||
Db 146 TTCGGGAAATCTCATTGACCGAGGGGGCCCTTTGGGCTATCAGAAGATCTCTGTGGATCCC 205

QY 1521 agcggagaccttgcggtccagattgccagctccggctcagctcgaactacactggcact 1580
      |||||
Db 206 TGC CGGGGGCTCCACCTTCACATTTCCCAATTCGACCTGCGACCGAGTCCCAATTACCTCGCAGCT 265

QY 1581 tcgtggccctggggcggtgccatcataatgggaactggctgtggtatccccctgggtc 1640
      |||||
Db 266 CCAGAGGCGCTGGGGCGCGCTCCATTATCAATGGGAACCTGGGCTGTGGATCCTCCAGSGTC 325

QY 1641 ctacagggcggggagaccttcttgataaaccttcctccaggaggagggcaagg 1700
      |||||
Db 326 CTATACAGCCATTGGGAGTCTCTTCAGATATACCGTCTCCAGGGAGCAAGCAAGG 385

QY 1701 ggagagtctgcggtgaagggcccccaaccagcctgtggatgtctatgatctttca 1760
      |||||
Db 386 GGAGAGTCTGTACGACGAAGGGCCCTACCACTCAGCCTGTGGAGCTCTATATGATCTTTCA 445

QY 1761 ggaggaaccagcgggtttttatcagtatgtcatctctcactcctccactcctga 1820
      |||||
Db 446 AGAGGACAAACCCAGGTGTTTTTATCAGTATGTCATCTTCCTCCTCGAGTCTTAGA 505

QY 1821 gaaccccaaccagagccccctgtccccagcttcagccggagattctgaggggtggagcc 1880
      |||||
Db 506 GAGTCTCTCCACAAAGCCTCCAGCCCTTTCAGCCTCAGCCTGAGCTGAGTGTGAGGGGGAGCC 565

QY 1881 ccacattgtccggcaccgcccgccagccagccagccagccacccctccagcgtcaggtgcg 1940
      |||||
Db 566 CTTACTCCCTCAGCCCCCGCCAGTCCGGGGGACACAGGCACCTCTCCAGCTCAGGTGCG 625

QY 1941 gatccccagatgcgcgcgcgcgcgcacatccagagacccccctggggtctccagctcgta 2000

```



Db 1 COTGGCTTCCGCTTCTATGTCCGTACACTGAAAGGTCCAGGATGGAGCCCTGTGTAG 60  
QY 1333 cctggagcccttgacatctgtgtgctggagcgtgtctgagcccccgcgtgtgatggatc 1392  
Db 61 GCTGGAGCCCTGACATCTGTGTGGCTGGAGCGTGTCTGAGCCCGCGCTGTGTGGATC 120  
QY 1393 attgagctgtgagcagcctcctgatggc-tgtgagagctgtgtgggggtgatattacatg 1451  
Db 121 CTTGGCTCTGGCAGCGCTCTGTATGGCTGTGGAGTCTGTGTGGGGTGTATGATTACCTG 180  
QY 1452 tgcctcttctcggggaacctcactgacccgagggggcccccctggctggtatcagagatctt 1511  
Db 181 TCGCCTTCTTTCGGGGAACCTCACTGACCGAGGGGGCCCTGGCTGGCTATCAGAGATCTT 240  
QY 1512 gtgattccagc-gggagccttgaggctccagattgcccagctccggcctagctccaaact 1570  
Db 241 GTGGATTCCAGCGGGAGCCCTTGGGCTGCCAGATTG-CCAGCTCCGGCCTAGCTCCAACT 299  
QY 1571 acctggaactctgtgaccttggggcc-ggtccatcatcaatgggaac---tgggctgtg 1626  
Db 300 ACCTGGCACTTCGTGGCCCTGGGGCCCTGGTCCATCATCAATGGGAACCTTGTGGCTTGTG 359  
QY 1627 gatccccctgggtctacagcggcggggacgtcttcttcgataataaaccgtcccccagg 1686  
Db 360 GATCCCTTGGCTCTACAGGCGCGGGACCGCTTCTTCGATATATAACCGTCCCGAGG 419  
QY 1687 g-aggagggcaaggaggagctgtgctggctgaaggcccccaccagcctgtggatgt 1745  
Db 420 GAAGGAGGCAAGGGAGAGTCTCGGCTGAAGG-CCCAACCAACAGCCCTGTGGATGT 478  
QY 1746 ctatatgtcttcaggaggaacccccggcggtttttatcagtatgtcatctcttccacc 1805  
Db 479 CTATATGATCTTTCAGGAGGAAACCCAGGCGTAGATAATCGAGTAGTGCCTCGAGTCTCT 538  
QY 1806 tctccaatccttgagaacccc 1827  
Db 539 GTACAGCGTCTGCTCGCAATACC 560

Query Match 15.6%; Score 411; DB 10; Length 437;  
Best Local Similarity 97.7%; Pred. No. 1.7e-74;  
Matches 417; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 tggactggcagccctggctgtatctgctgctgtctcttctccctccctcagctctgcttg 69  
Db 7 TTGTTTTCAGGCGCCCTGGCTGTATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 66  
QY 70 gatcaggaggtgtgttcgggacactctcttcagacacacacacagagagggccagggcccc 129  
Db 67 GATCAGGAGGTGTGTCCGGGACACTCTCTTCAGACACCTACAGAGGAGGGCCAGGGCCCC 126  
QY 130 gaagggtgtctgggggaccttgggtccagtggtggtctcttctccagccctgcgggggtgggg 189  
Db 127 GAAGGTGTCTGGGGACCTTGGGTCCAGTGGGCTCTGTCTTCCAGCCCTGCGGGGTGGGG 186  
QY 190 gtgcagcgcagggagcgggacatgtcagctccctacagtcagctccacccagctctgccc 249  
Db 187 GTGCAGCGCAGGAGCGGACATGTCTAGCTCTCTTACAGTGAGCTCCACCGAGTCTGCC 246  
QY 250 ctccctcccgcccccacagacatccagagccctctcccccgggggaggggtccacaga 309  
Db 247 CTCCTCCCGCGCCCGCCCAAGACATCCAGAGCCCTCTCTCCCGCGGGGCCAGGTTCCCGA 306  
QY 310 cccagacttctccagaaacccctccctgtacagagacacagctctcggggaagggtggc 369  
Db 307 CCCAGACTTCTCCAGAAACCCCTCCCTTGTACAGGACACAGTCTCGGGGAAGGGGTGGC 366  
QY 370 ccacttcgaggtcccgcttccaccctaggagagagagagagagagagagagagagagag 429  
Db 367 CCATTCAAGAGCCCGCTTCCACCTAGGAGAGAGAGAGAGAGAGAGAGATTCAGAGCGGAC 426  
QY 430 aggaagt 436  
Db 427 AGCAGGT 433

BASE COUNT 76 a 151 c 131 g  
ORIGIN

RESULT 6  
BE808908  
LOCUS 214215 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001  
DEFINITION  
ACCESSION BE808908  
VERSION BE808908.1 GI:10240020  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 517)

AUTHORS	TITLE	BASE COUNT
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G., Perteira, G., Holt, I.I., Karanycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle	84 a 175 c 179 g 79 t
21180013	Genome Res. 11 (4), 626-630 (2001)	
Contact: Smith TPL		
USDA, ARS, US Meat Animal Research Center		
PO Box 166, Clay Center, NE 68933-0166, USA		
Tel: 402 762 4366		
Fax: 402 762 4390		
Email: smith@email.marc.usda.gov		
Single pass sequencing, Bases called and alt.trimmed with phred v0.980904.e. vector identified by cross_match with the -mscore 18 and -minmatch 12 options.		
PCR Primers		
FORWARD: AGGAAACAGCTATGACCAT		
BACKWARD: GTTTCGCCAGTCACGACG		
Plate: 70 row: N column: 6		
Seq primer: ATTAGGTGCACACTATAG.		
Location/Qualifiers		
1.517		
/organism="Bos taurus"		
/db_xref="taxon:9913"		
/clone_lib="MARC 2BOV"		
/tissue_type="pooled"		
/lab_host="DH10B"		
/note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."		
FEATURES		
SOURCE		
ORIGIN		

Query Match	14.9%	Score 391.4	DB 10	Length 517
Best Local Similarity	87.5%	Pred. No. 2e-70		
Matches	428	Conservative	0	Mismatches 61
				Indels 0
				Gaps 0
Qy	2071	tgatctcccgctgagtcgggagagaaatgataaacagagctgtgcgcgggtgccagg	2130	
Db	1	TGCATTTCGAGAGTCAGAGAGAGCTGGATGAACGCAGCTGTGCATTGGGTGCCAGG	60	
Qy	2131	ccccagagctccccgtgaacctgccacggaccacccatgcccccatactggggaggctggc	2190	
Db	61	CCCCAGCCTCCAGSAGGCCCTGCCACGGCCCCCGTGCACCATACTGGGAGGCGGC	120	
Qy	2191	gagtgcacatctcagacogctctctgtgcccgcagccacagacgcagctgcagtcg	2250	
Db	121	GAGTGGACGCTCTGCGAGCCGTTGCTGTGGACCGCGCACCCACGACACCTCAGCTACGCTGC	180	
Qy	2251	cggcaggaattggggggggtggctctctcgtgtcccccgagcgtgtgacatatcccc	2310	
Db	181	CGGCAGGAGTTTGGGGGTGGCGGCTCTCAGTGCOCCTAGAGCGCTCGCGGCACCTGCC	240	
Qy	2311	cggcccaacatcacccagcttgcagctgcgctctgtgcccattgggaagtggctct	2370	
Db	241	CGACCCAAACATCACCCAGGCCCTGCCAGCTGCCTCTGTGGCCCATTTGGGAGGTTGCTCA	300	
Qy	2371	ccttgagcagtgctccgtgcggtgcgcccggggccagagaaacggcgaggttcgctgt	2430	
Db	301	CCCTGGAGTCAGTGCTCTCTGGCGATCGCGGCGCGGTCTCAGAGGAGCCGCGCAGCTCCGCTGT	360	
Qy	2431	gttgggaacaacggtataaagtgcagacagagaggttgcgtcagccccccacagccc	2490	
Db	361	GTCGGCAACACCGGGATGAAGTGAGTGAACAGAGAGTGCCTCATGGCCCCCGGGGCC	420	
Qy	2491	ccccagagagcctgtgacatggggccctgtactactgctggttccacagagctgg	2550	
Db	421	CCCAGCAGAGAGGCTTGCACATTGGGGCCCTGCACACAGGGCTTGTTTCCACAGGACTGG	480	

QY	2551 agctccaag	2559       
Dd	481 AGCTCCAAG	489
RESULT	7	
AW230557		
LOCUS	awc65all.y1 NCI_CGAP MamI Mus mRNA linear EST 10-DEC-1999 IMAGE:2647388 5'	
DEFINITION	similar to TR:O60345 O60345 KIAA0605 PROTEIN.; Muridae; Murinae; Mus.	
ACCESSION	AW230557	
VERSION	AW230557.1 GI:6559853	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1. (bases 1 to 495)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Other_ESTs: uc65all.xl Contact: Robert Strausberg, Ph.D. Email: cgapsr@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="#">www.bio.lnlni.gov/bbrp/image/image.html</a>	
MGI:1027840		
Seq primer:	-40RP from Gibco	
High quality sequence stop:	357.	
Location/Qualifiers		
FEATURES		
source	1..495	
/organism=	"Mus musculus"	
/strain=	"FVB/N"	
/db_xref=taxon:	10090"	
/clone=	"IMAGE:2647388"	
/clone.lib=	"NCI_CGAP_MamI"	
/tissue_type=	"tumor, biopsy sample"	
/dev_stage=	"10 months, virgin"	
/lab_host=	"DH10B"	
/note=	"Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;	
Site 2:	NotI; Cloned unidirectionally. Primer: Oligo di-	
Library constructed by:	Life Technologies. Investigator	
providing samples:	Gilbert Smith, NIH	
a	155 c	149 q
b	155 c	149 g
c	155 c	149 t
d	155 c	149 t
e	155 c	149 t
f	155 c	149 t
g	155 c	149 t
h	155 c	149 t
i	155 c	149 t
j	155 c	149 t
k	155 c	149 t
l	155 c	149 t
m	155 c	149 t
n	155 c	149 t
o	155 c	149 t
p	155 c	149 t
q	155 c	149 t
r	155 c	149 t
s	155 c	149 t
t	155 c	149 t
BASE COUNT		

	Query Match	14.2%	Score 373.4	DB 9	Length 495
	Best Local Similarity	84.6%	Prod. No. 9.9e-67		
	Matches 419	Conservative 0	Mismatches 76	Indels 0	Gaps 0
QY	1939	cggatccccccagatgcgcgcgcgcgcgcacatccacagacacccctcggggtctccagctgcg	1998		
Db	1	CGGATTCCCAAGTCTCTCCGACTCTGTCTCAGACAGCCATGGGCTCTTCGTCTGGA	60		
QY	1999	tactggaacgagtgaggacaactctcatctcagcgtccctgcgggaaagtgctctggcgc	2058		
Db	61	TACTGGAAACAAGTGGGGCACTCTGAGTGTTCAGCATCTGTGGCAAGGTGTTGGCAC	120		
QY	2059	ccattttctctgcattcccgtagctcggagaggaactgtagaagcagctgtgc	2118		
Db	121	CCATTTTCTCTTGATTTCCCGTGNCTCAGAGAGGAGTTGGATGAACAGAGCTTGCT	180		
QY	2119	gcgggtgccaggccccccagctccctctgaacctgcacaggcaccoccatgcccccatac	2178		

QY	1416	tgctgtgagatctgtggggtgatgtattacattcgtgcctgttctt-ggggaacctca	1474
Db	2	TGGCTGTGAGTCTGTGGGGTGATGATCTACCTGTGCCTTGTCTTGGGGAACCTCA	61
QY	1475	ctgaccgaagggccccctggcctatcagaagatcttgatctcacgagccttgc	1534
Db	62	CTGACCGAGGGGGCCCCCTGGGCTATCAGAAGATCTTGGATTCACGGGAGCCTTGC	121
QY	1535	ggctccagattgccagctccggcctagctccaaactcctggcaattcgtggccctggg	1594
Db	122	GGCTCCAGATTGCCAGCTCCGGCTAGTCTCCAACTACCTGGCACTTGTGGCCCTGGGG	181
QY	1595	gccgttccatcatatgggaactgggctgtgatcccccctgggtccctacaggccggcg	1654
Db	182	GCGGFPCCATCATCAATGGGAAGTGGCTGTGGATCCCCCTGGTCCATACAGGGCCGGC	241
QY	1655	ggaccgtctttcgatataaccgtccctccag-ggaggagggcaaaaggggagagtctgc	1713
Db	242	GGACCGCTCTTCGATATAACCGTCTCCAGTGGAGGAGGCAAGGGGAGAGTCTGTGC	301
QY	1714	gtgaaggcccccacacccagcctgtggatgctatgatctttcaggaggaaaccca	1773
Db	302	GCTGAAGGGCCCAACACCCAGGCTGGGATGCTATATGATCTTCAGGAGAAACCCA	361
QY	1774	ggcgtttttatcagtatcatctcttccactccctcccaatccttgagaa	1823
Db	362	GGCG-TTTTTATCAGTATGTCATCTCTTCACTCTCCATCTCCATCTTGAGGA	410
RESULT	9		
BI330332		866 bp	linear
LOCUS	602983024F1	NCI_CGAP_L19	Mus musculus cDNA clone IMAGE:5135860 5',
DEFINITION		mRNA	
ACCESSION	BI330332		
VERSION	BI330332.1	GI:15014989	
KEYWORDS	EST.		
SOURCE		house mouse.	
ORGANISM		Mus musculus	
REFERENCE		Ekuryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
TITLE		1 (bases 1 to 866)	
JOURNAL		NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC)	
		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgabp@remail.nih.gov	
		Tissue Procurement: Jeffrey E. Green, M.D.	
		CDNA Library Preparation: Life Technologies, Inc.	
		DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LLNL at:	
		<a href="http://image.llnl.gov">http://image.llnl.gov</a>	
		Plate: LLAM1333 row: p column: 05	
		High quality sequence start: 3	
		High quality sequence stop: 540.	
FEATURES		Location/Qualifiers	
source		1..866	
		/organism="Mus musculus"	
		/strain="FVB/N"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:5135860"	
		/clone_lib="NCI_CGAP_L19"	
		/lab_host="DH10B (T1 phage-resistant)"	
		/note="Organ: liver; Vector: pCMV-SpOrn6; Site:1: NotI;	
		Site:2: SalI; cloned unidirectionally. Primer: oligo dr.	
		Average insert size 1.9 kb. Constructed by Life	
		Technologies. Note: this is a NCI_CGAP Library."	
BASE COUNT	149 a	332 c	237 g 148 t
ORIGIN			

TITLE	JOURNAL	MEDLINE	COMMENT
Brustein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&amp;tt2=IL5-GN0178-311000-204.f10&amp;t3=2000-10-31&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&amp;tt2=IL5-GN0178-311000-204.f10&amp;t3=2000-10-31&amp;t4=1</a> ) Seq primer: puc 18 forward High quality sequence stop: 305. Location/Qualifiers 1..360 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="GN0178" /dev_stage="Adult" /note="Organ: placenta_normal; Vector: puc18; Site_1: Sma1 ; Site_2: Sma1; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 59 a 80 c 142 q 78 t 1 others			

```

BASE COUNT      59 a      80 c      142 g      78 t      1 others
ORIGIN

Query Match      13.6%   Score 357.4;   DB 10;   Length 360;
Best Local Similarity 99.4%;   Pred. No. 1.7e-63;
Matches 358;   Conservative      0;   Mismatches 2;   Indels 0;   Gaps 0;

```

QY	470	tcggtatggagagtgocctttgcatcgcacgcacccggaacccgagggacacctcgga	529
Db	360	TCGGTTATGGGAGAGTGCCTTNTGATTCGCACATGCACCGGAACCGCAGGACACCTCGGA	301
QY	530	gccacaccagatcgagctgcctctgactctctcctagaggggaagagctattccgtccc	589
Db	300	GCCACCCAGATCTGAGCTGTCCCTGATCTCTTCTAGAGGGGAAGAGCCTATTCCGTCCTC	241
QY	590	ctactccaaagcagagcccatctctccgcaaacgcagcccccacaaactcagctccctccca	649
Db	240	CTACTCCAAGAGCAGAGCCATTCTCCGCAACGCGACGCCCAACTGAGTCCCTCCCA	181
QY	650	cagaaactgtctgtccacacccccatcccccccaagcagaaacctctaagccctgaaaactgtcc	709
Db	180	CAGAACTGTCTCTCCACACCCCATCCCCCCCAAGCAGAACCTCTTAAGCCCTGAAACTGTCTC	121
QY	710	agacagaggtggccccccagaacagagcgtgccccctcaggcatcacccagagcccaag	769
Db	120	AGACAGAGGTGGCCCCCAGAACCCAGGCGTGCCCCCTACGGCATCACCCAGAGCCCCAAG	61
QY	770	cctctggcacagagccccccctcacccacgcactccttagggaaggtggcttcttccgtg	829
Db	60	CCTCTGGCACAGAGCCGCCCTCACCCAGCAGCTCCTTAGGAAAGGTGGCTTCTTCCGTG	1

RESULT	11
BM391602	
LOCUS	BM391602 592 bp mRNA linear EST 17-JAN-2002
DEFINITION	UI-R-DY0-ckr-l-07-0-UI.s1 UI-R-DY0 Rattus norvegicus CDNA clone
ACCESSION	UI-R-DY0-ckr-l-07-0-UI 3', mRNA sequence.
	BM391602

Query Match	13.7%;	Score 359.8;	DB 10;	Length 866;
Best Local Similarity	76.6%;	Pred. No. 7.6e-64;		
Matches 557;	Conservative	0;	Mismatches 157;	Indels 13; Gaps 9;
1125	ccaaagcgcctgccccctgagcagcagaccccccggccctgcagtcgagcgccttttaa	1184		
1	CCAGGAGCCCTTGCCCCCTGAGCATGCAGACCCAGGCCCTACAGTGTGCGCGCTTTGA	60		
1185	ctcccaggaattcatagggccagctgtatcagtcggagcccttcaactcaagtccagggtc	1244		
61	CTCCCAGGAATTATGGCGCAGCTGTACAGTGGAGGCCCTTACCGAAGTTCAGGGCTC	120		
1245	ccagcgcgtgaaactgaactgcgcgcccccgctggcttccgcttctatgtccgtcacactga	1304		
121	CCAGCGCTGTGAATGAACCTGACCTGCGCCCGCGTCCGATTCTATGCCACACACTGA	180		
1305	aaagctccagatagggaacctgtgtcagctcagcagagccctgaactctgtgtgcctgagc	1364		
181	AAAGTGCAGATGGAACCCCTGTGTGAGCTGTGATCCCTAGACATCTGTGTGGCTGGAGC	240		
1365	ctgctcagcccgctgtgatggatcttggctcgtgcagcgcctcctg-atggctgtg	1423		
241	CTGCCTGAGCCCTGGCTGTGATGGGTCTTGCTCTGGCAGGCGTCGGGCAATGGCTCG	300		
1424	gagctctgggggtga-tgaattcaactctgccttgtttctggg-gaacctcaactgacgc	1481		
301	GAGTCTCGGGGGTGACTGGTTCTACTCGCGCTGGTTCCGGGCAAACTCACATGACCG	360		
1482	agggggccccctgggctatcagaa--gatcttggattccagcgggagccttgcgctc	1539		
361	AGGGGCCCTTGGGCTATCAGATCGATCTGTGATCCCTGCCGGGCCCTCCACACTT	420		
1540	cagattgccagctccggctagctcaactaactgcactctgtgcctcgtggggcgcg	1599		
421	CACATTCCCAAGTGGCAGCC-AGTTCAAATTACTCGCACTCCGAGGGCTCGGGGCGC	479		
1600	tcca-tcaatcaatgggaactggcgtggtacccccctggg-tctcacaggccgagcagga	1657		
480	TCCATTATCATACTGGGAACCTGGGCTGTGGATCTCCAGGGCTCTCTACAGCATCGCGA	539		
1658	ccgtctttogatataaacgctctctcccaagg---gagagggcaagggagagctgtctg	1713		
540	CTGTCTTCCAGTATAAGCGTCTCCAGGGCAAGACAGCGCCAGCGGGAGAGTGTGCA	599		
1714	gctgaaggccccacaccccgctgtg-gatgtctatatgatctttcaggagagaaacc	1772		
600	GCCAAAGGCCCTTACCATCGACCTGGCGCGCTCTCTGTGACCCACCCACCGAGGACCC	659		
1773	agggctttttatcagtatgtcatctcttcaactctcaactcttgagaacccccccc	1832		
660	CCCCAGGGTGTGCCCGACCCCGCATGCTCATCCACCCCCCTCCCGGAGCGCCGACAAG	719		
1833	agagccc	1839		
720	AGGCC	726		
RESULT 10	BF993048/c			
LOCUS	BF993048	360 bp	mRNA	linear
DEFINITION	I15-GN0178-311000-204-f10 GN0178 Homo sapiens cDNA, mRNA sequence.			
ACCESSION	BF993048			
VERSION	BF993048.1	GI:12399371		
KEYWORDS	EST			

SOURCE	ORGANISM	Human	Non sapiens	Other
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
	1. (bases 1 to 360)			
REFERENCE	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zagó, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.			
AUTHORS				



QY	2398	ggccggggccagagaagccggcaggttcgtgtgttggggaacaacagtgatgaagtgcagc	2457
DB	331	ggtctgtgttcagagagagccgcaagtggtgtgttgaagtaataatggccatgagtggtgc	390
QY	2458	gagcaggagtggtggttcaggcccccacagccccccagcagagagcgtgtgacatgggg	2517
DB	391	aaagcaggagtggtggttcaggcccccacagccccccagcagagagcgtgtgacatgggc	450
QY	2518	ccctgtactactgctgtgttcacagagcagcgtgtgacatgggc	2559
DB	451	ccctgtactactgctgtgttcacagagcagcgtgtgacatgggc	492
RESULT	12		
LOCUS	BG005397	370 bp	linear
DEFINITION	PM1-GN0180-241100-003-e05 GN0180 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BG005397		
VERSION	BG005397.1	GI:12447519	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 370)		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed		
JOURNAL	sequence tags		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
COMMENT	20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM1&t2=PM1-GN0180-241100-003-e05&t3=2000-11-24&t4=1) Seq primer: puc 18 forward High quality sequence start: 7 High quality sequence stop: 370.		
FEATURES	Location/Qualifiers		
source	1..370 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="GN0180" /dev_stage="Adult" /note="Organ: placenta normal; Vector: puc18; Site: 1; SmaI ; Site: 2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." BASE COUNT 80 a 127 c 98 g 64 t 1 others ORIGIN		
Query Match	12.8%; Score 337; DB 10; Length 370;		
Best Local Similarity	99.4%; Pred. No. 2.8e-59;		
Matches	348; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		
QY	322	ccagaaacccctcccttctgtac-agcacagctctcgggaagggtgcccattcagc	380



```
High quality sequence stop: 375.
FEATURES
  source
    1..375
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="GN0187"
        /dev_stage="Adult"
        /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
        products derived from ORESTES PCR (U.S. Letters Patent
        application No. 196,716 - Ludwig Institute for Cancer
        Research) profiles into the pUC 18 vector. Reverse
        transcription of tissue mRNA and cDNA amplification were
        performed under low stringency conditions."
BASE COUNT      81 a 124 c 106 g 63 t 1 others
ORIGIN
Query Match      12.1%; Score 317.8; DB 10; Length 375;
Best Local Similarity 97.4%; Pred. No. 2.5e-54;
Matches 333; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 312 ccagactctccagaaacccctcccttctacagacacagctcgcgggaagggtggccc 371
Db 27 CCAGACTTCTCCAGAAACCTCCCTCC-TCTACAGGACACAGTCTCGGGGAAGGGTGGGCC 85
QY 372 acttcagaggtccggttcccaactaggagagagagagaccagagattcagcggccag 431
Db 86 ACTTCAGGTCCCGTCTCCACCTAGGAGAGAGAGACCAGAGATTTCAGCGGCCAG 145
QY 432 gagggtcccggttcagaccctccatcaagccaggaatgttcggttatggagagtgccctt 491
Db 146 GAGGTCCCGTTCAGACCCCTCATCAGCCAGGAATGTTCGGTATGGAGAGTGCCCTT 205
QY 492 tgcattgcacatgcacccggaacccgagcaccctcggagccaccacagatctgagctgtc 551
Db 206 GCGATGGCCACGGCACCAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 265
QY 552 cctgatctcttcagagggagagctatccgtccctcactcaagagagagagcatt 611
Db 266 CTGATCTCTTCTAGGGGAGAGACCTATTCTGTCCCTCTACTCCAGAGCAGAGCCATT 325
QY 612 ctccgaaacgagcagcccccacaaactgagctccctcccccacaga 653
Db 326 CTCGCAAGCGCAGCGCCCAANTGAGTCCGCCCCACAGA 367

RESULT 15
BF993027
LOCUS      BF993027      324 bp      mRNA      linear      EST 23-JAN-2001
DEFINITION IL5-GN0178-311000-204-e09 GN0178 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF993027
VERSION     BF993027.1  GI:12399350
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 324)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nazai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
```

```
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-GN0178-
311000-204-e09&t3=2000-10-31&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 315.
FEATURES
  source
    1..324
      Location/Qualifiers
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="GN0178"
        /dev_stage="Adult"
        /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
        products derived from ORESTES PCR (U.S. Letters Patent
        application No. 196,716 - Ludwig Institute for Cancer
        Research) profiles into the pUC 18 vector. Reverse
        transcription of tissue mRNA and cDNA amplification were
        performed under low stringency conditions."
BASE COUNT      69 a 115 c 85 g 55 t
ORIGIN
Query Match      11.9%; Score 314.4; DB 10; Length 324;
Best Local Similarity 99.7%; Pred. No. 1.2e-54;
Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 363 ggtggccacttcgaggtccgccttccacactagagagagagagaccagagattcg 422
Db 1 GGTGGCCCACTTCGAGGTCCCGCTTCGAGACCCCATCAAGCCAGGAATGTTCGGTTATGGGAG 60
QY 423 agcgccagaggttcgcggttcgagaccctcaagccaggaatgttcggttatggag 482
Db 61 AGCGGCCAGGAGGTCCCGGCTTCGAGACCCCATCAAGCCAGGAATGTTCGGTTATGGGAG 120
QY 483 agtgcctttgcattccactgcacggaaacccagcaccctcgagccaccagatc 542
Db 121 AGTCCCTTTTGCAATGCCACTGCACCGGAACCCAGCACCCCTCGGAGGCCCCAGATC 180
QY 543 tgagctgcctctgatctcttctagagggggaagggtatttcctccctactccaagagc 602
Db 181 TGAGCTGTCCCTGATCTCTTCTAGAGGGGAAGAGCCTATTCCGTCCTCCCTACTCCAAGAGC 240
QY 603 agagccatttcgcaaacggcagcccccacaaactgagctccctccacagaaactgtctgt 662
Db 241 AGAGCCATTCTCGCAACCGCAGCCCCCAACTGAGCTCCCTCCACAGAACTGTCGT 300
QY 663 ccacacccccatcccc 678
Db 301 CCACACCCCATCCCC 316

Search completed: July 23, 2002, 16:20:47
Job time: 11957 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 18:22:05 ; Search time 64.95 Seconds  
(without alignments)  
9961.488 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1 atgggaactggactggcag.....cctcagccttcacgcatag 2634

Scoring table: OLIGO.NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	ID		
1	19	0.7	410	3	US-09-114-146-1	Sequence 1, Appli
2	19	0.7	410	5	PCT-US96-08623-1	Sequence 1, Appli
3	19	0.7	2055	3	US-08-872-855-3	Sequence 3, Appli
4	19	0.7	2800	3	US-08-872-855-1	Sequence 1, Appli
5	18	0.7	1383	1	US-08-484-044-1	Sequence 1, Appli
6	18	0.7	1787	2	US-08-372-652-6	Sequence 6, Appli
7	18	0.7	1787	5	PCT-US95-16311-6	Sequence 6, Appli
8	18	0.7	2511	2	US-08-422-699A-8	Sequence 8, Appli
9	18	0.7	2511	2	US-08-422-706B-8	Sequence 8, Appli
10	18	0.7	2607	2	US-08-907-156-1	Sequence 1, Appli
11	18	0.7	2610	2	US-08-989-386-2	Sequence 2, Appli
12	18	0.7	2726	2	US-08-422-699A-12	Sequence 12, Appli
13	18	0.7	2726	2	US-08-422-706B-12	Sequence 12, Appli
14	18	0.7	3182	1	US-08-484-044-11	Sequence 11, Appli
15	18	0.7	3323	2	US-08-422-699A-10	Sequence 10, Appli
16	18	0.7	3323	2	US-08-422-706B-10	Sequence 10, Appli
17	18	0.7	4508	5	PCT-US93-06251-34	Sequence 34, Appli
18	18	0.7	11613	1	US-08-484-044-10	Sequence 10, Appli
19	17	0.6	300	3	US-09-157-177-117	Sequence 117, App
20	17	0.6	427	2	US-08-623-906A-5	Sequence 5, Appli
21	17	0.6	522	1	US-08-428-732-1	Sequence 1, Appli
22	17	0.6	636	4	US-08-930-117-7	Sequence 7, Appli
23	17	0.6	812	1	US-08-656-253-1	Sequence 1, Appli
24	17	0.6	1581	1	US-08-313-553-8	Sequence 8, Appli
25	17	0.6	1581	3	US-08-767-993-8	Sequence 8, Appli
26	17	0.6	1683	1	US-08-164-614A-5	Sequence 5, Appli
27	17	0.6	1683	2	US-08-456-489B-5	Sequence 5, Appli

C 28	17	0.6	1683	5	PCT-US93-01720-5	Sequence 5, Appli
C 29	17	0.6	1758	1	US-08-078-222B-1	Sequence 1, Appli
C 30	17	0.6	1758	2	US-08-661-330A-1	Sequence 1, Appli
C 31	17	0.6	1758	3	US-09-038-217A-1	Sequence 1, Appli
C 32	17	0.6	1810	1	US-08-293-849B-20	Sequence 20, Appli
C 33	17	0.6	1810	2	US-08-142-388A-20	Sequence 20, Appli
C 34	17	0.6	1810	3	US-08-967-727-20	Sequence 20, Appli
C 35	17	0.6	1810	4	US-08-037-230D-20	Sequence 20, Appli
C 36	17	0.6	1892	3	US-08-620-077B-2	Sequence 4, Appli
C 37	17	0.6	1947	1	US-08-164-614A-4	Sequence 4, Appli
C 38	17	0.6	1947	2	US-08-456-489B-4	Sequence 4, Appli
C 39	17	0.6	1947	5	PCT-US93-01720-4	Sequence 4, Appli
C 40	17	0.6	1956	1	US-08-313-553-6	Sequence 6, Appli
C 41	17	0.6	1956	3	US-08-767-993-6	Sequence 6, Appli
C 42	17	0.6	1997	1	US-08-164-614A-6	Sequence 6, Appli
C 43	17	0.6	1997	2	US-08-456-489B-6	Sequence 6, Appli
C 44	17	0.6	1997	5	PCT-US93-01720-6	Sequence 6, Appli
C 45	17	0.6	2040	3	US-08-604-789B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-09-114-146-1  
; Sequence 1, Application US/09114146  
; Patent No. 6083747  
; GENERAL INFORMATION:  
; APPLICANT: WONG, Peter M.C.  
; APPLICANT: CHUNG, Siu-Wah  
; APPLICANT: HAN, Xiaodong  
; TITLE OF INVENTION: GLYCOPROTEIN GP105 ON BLS HEMATOPOIETIC  
; TITLE OF INVENTION: STEM CELLS  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/114,146  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/471,188  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 46074/102/FEIN  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-114-146-1

Query Match 0.7%; Score 19; DB 3; Length 410;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1417 ggcgtgagtgctgtggg 1435  
|||||  
Db 340 GGCCTGGAGTCTGTGGG 358

RESULT 2  
PCT-US96-08623-1  
; Sequence 1, Application PC/TUS9608623  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: GLYCOPROTEIN GP105 ON BLS HEMATOPOIETIC  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08623  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/471,188  
; FILING DATE: 06-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 46074/103/FEIN  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; PCT-US96-08623-1

Query Match 0.7%; Score 19; DB 5; Length 410;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1417 ggcgtgagtgctgtggg 1435  
|||||  
Db 340 GGCCTGGAGTCTGTGGG 358

RESULT 3  
US-08-872-855-3/c  
; Sequence 3, Application US/08872855  
; Patent No. 6121045  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean  
; APPLICANT: Gearing, David  
; TITLE OF INVENTION: NOVEL HUMAN DELTA3 COMPOSITIONS AND  
; TITLE OF INVENTION: THERAPEUTIC USES THEREFOR  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA

; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/872,855  
; FILING DATE: 11-JUN-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: MAA-003.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2055 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-872-855-3

Query Match 0.7%; Score 19; DB 3; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2404 gccacagagccgcgag 2422  
|||||  
Db 1417 GCCACAGAGCGGCAGG 1399

RESULT 4  
US-08-872-855-1/c  
; Sequence 1, Application US/08872855  
; Patent No. 6121045  
; GENERAL INFORMATION:  
; APPLICANT: McCarthy, Sean  
; APPLICANT: Gearing, David  
; TITLE OF INVENTION: NOVEL HUMAN DELTA3 COMPOSITIONS AND  
; TITLE OF INVENTION: THERAPEUTIC USES THEREFOR  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/872,855  
; FILING DATE: 11-JUN-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: MAA-003.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2800 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338..2392
; US-08-872-855-1

Query Match 0.7%; Score 19; DB 3; Length 2800;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2404 gccacagagacgcgcagc 2422
Db 1754 gccacagagacgcgcagc 1736

RESULT 5
US-08-484-044-1/c
; Sequence 1, Application US/08484044
; Patent No. 5552282
; GENERAL INFORMATION:
; APPLICANT: Caskey, C. T.
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Friedman, David L.
; APPLICANT: Pizzuti, Antonio
; APPLICANT: Fenwick, Raymond G.
; TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,044
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/019,940
; FILING DATE: 19-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1383 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-484-044-1

Query Match 0.7%; Score 18; DB 1; Length 1383;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 245 gccctccctcccccgc 263
```

```
Db 1050 gccctccctcccccgc 1033

RESULT 6
US-08-372-652-6/c
; Sequence 6, Application US/08372652
; Patent No. 5932699
; GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Seol, Wongi
; APPLICANT: Choi, Hwang-Sik
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
; POLYPEPTIDES AND RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street, Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/372,652
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/246001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-372-652-6

Query Match 0.7%; Score 18; DB 2; Length 1787;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 690 tctaagccctgaaactgc 707
Db 37 tctaagccctgaaactgc 20

RESULT 7
PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311
; GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Seol, Wongi
; APPLICANT: Choi, Hwang-Sik
; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
; POLYPEPTIDES AND RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street, Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
```

```

; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16311
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/372,652
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/246001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1787 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; PCT-US95-16311-6

Query Match 0.7%; Score 18; DB 5; Length 1787;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 690 tctaaagccctgaactgc 707
Db 37 TCTAAGCCCTGAAACTGC 20

RESULT 8
US-08-422-699A-8/C
; Sequence 8, Application US/08422699A
; Patent No. 5955265
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,699A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,706
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/023,612

```

```

; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/839,255
; FILING DATE: 20-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01545
; FILING DATE: 19-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00253
; FILING DATE: 05-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9202485.0
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5830A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1746
; US-08-422-699A-8

Query Match 0.7%; Score 18; DB 2; Length 2511;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccgccc 263
Db 1920 GCCCTCCCTCCCGGCC 1903

RESULT 9
US-08-422-706B-8/C
; Sequence 8, Application US/08422706B
; Patent No. 5977333
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,706B
; FILING DATE: 14-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/284,543

```

;> FILING DATE: 08-AUG-1994  
;> PRIOR APPLICATION DATA: US 08/023,612  
;> FILING DATE: 26-FEB-1993  
;> PRIOR APPLICATION DATA:  
;> APPLICATION NUMBER: US 07/839,255  
;> FILING DATE: 20-FEB-1992  
;> PRIOR APPLICATION DATA:  
;> APPLICATION NUMBER: PCT/US93/01545  
;> FILING DATE: 19-FEB-1993  
;> PRIOR APPLICATION DATA: PCT/GB93/00253  
;> APPLICATION NUMBER: PCT/GB93/00253  
;> FILING DATE: 05-FEB-1993  
;> PRIOR APPLICATION DATA:  
;> APPLICATION NUMBER: GB9202485.0  
;> FILING DATE: 06-FEB-1992  
;> ATTORNEY/AGENT INFORMATION:  
;> NAME: Granahan, Patricia  
;> REGISTRATION NUMBER: 32,227  
;> REFERENCE/DOCKET NUMBER: MIT-5830A2  
;> TELECOMMUNICATION INFORMATION:  
;> TELEPHONE: 617-861-6240  
;> TELEFAX: 617-861-9540  
;> INFORMATION FOR SEQ ID NO: 8:  
;> LENGTH: 2511 base pairs  
;> TYPE: nucleic acid  
;> STRANDEDNESS: double  
;> TOPOLOGY: linear  
;> MOLECULE TYPE: DNA (genomic)  
;> FEATURE:  
;> NAME/KEY: CDS  
;> LOCATION: 1..1746  
;> US-08-422-706B-8

Query Match 0.7%; Score 18; DB 2; Length 2511;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 gccctccctccctccggcc 263  
|||||  
Db 1920 gccctccctccctccggcc 1903

RESULT 10  
US-08-907-166-1/c  
;> Sequence 1, Application US/08907166  
;> Patent No. 5948666  
;> GENERAL INFORMATION:  
;> APPLICANT: Callen, Walter  
;> APPLICANT: Mather, Eric  
;> TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES  
;> FILE REFERENCE: 09010/027001  
;> CURRENT APPLICATION NUMBER: US/08/907,166  
;> CURRENT FILING DATE: 1997-08-06  
;> NUMBER OF SEQ ID NOS: 12  
;> SOFTWARE: FastSEQ for Windows Version 3.0  
;> SEQ ID NO 1  
;> LENGTH: 2607  
;> TYPE: DNA  
;> ORGANISM: Ammonifex degensii  
;> FEATURE:  
;> NAME/KEY: CDS  
;> LOCATION: (1)...(2604)  
;> US-08-907-166-1

Query Match 0.7%; Score 18; DB 2; Length 2607;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 834 cccctcagccagcaggcc 851

Db 1892 CCCTCAGCCACGAGGCC 1875  
|||||

RESULT 11  
US-08-989-386-2/c  
;> Sequence 2, Application US/08989386  
;> Patent No. 5989860  
;> GENERAL INFORMATION:  
;> APPLICANT: Bandman, Olga  
;> APPLICANT: Hillman, Jennifer L.  
;> APPLICANT: Guegler, Neil J.  
;> APPLICANT: Corley, Neil C.  
;> APPLICANT: Shah, Purvi  
;> TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS  
;> NUMBER OF SEQUENCES: 9  
;> CORRESPONDENCE ADDRESS:  
;> ADDRESSEE: Incyte Pharmaceuticals, Inc.  
;> STREET: 3174 Porter Drive  
;> CITY: Palo Alto  
;> STATE: CA  
;> COUNTRY: USA  
;> ZIP: 94304  
;> COMPUTER READABLE FORM:  
;> MEDIUM TYPE: Diskette  
;> COMPUTER: IBM Compatible  
;> OPERATING SYSTEM: DOS  
;> SOFTWARE: FastSEQ for Windows Version 2.0  
;> CURRENT APPLICATION DATA:  
;> APPLICATION NUMBER: US/08/989,386  
;> FILING DATE:  
;> CLASSIFICATION:  
;> PRIOR APPLICATION DATA:  
;> APPLICATION NUMBER:  
;> FILING DATE:  
;> ATTORNEY/AGENT INFORMATION:  
;> NAME: Billings, Lucy J  
;> REGISTRATION NUMBER: 36,749  
;> REFERENCE/DOCKET NUMBER: PF-0443 US  
;> TELECOMMUNICATION INFORMATION:  
;> TELEPHONE: 650-853-0555  
;> TELEFAX: 650-845-4166  
;> TELEX:  
;> INFORMATION FOR SEQ ID NO: 2:  
;> SEQUENCE CHARACTERISTICS:  
;> LENGTH: 2610 base pairs  
;> TYPE: nucleic acid  
;> STRANDEDNESS: single  
;> TOPOLOGY: linear  
;> IMMEDIATE SOURCE:  
;> LIBRARY: OVARTUT01  
;> CLONE: 2255114  
;> US-08-989-386-2

Query Match 0.7%; Score 18; DB 2; Length 2610;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1684 agggaggaggcagg 1701  
|||||

Db 2323 AGGAGAGGGGCAAGGG 2306

RESULT 12  
US-08-422-699A-12/c  
;> Sequence 12, Application US/08422699A  
;> Patent No. 5955265  
;> GENERAL INFORMATION:  
;> APPLICANT: Brook, J. David  
;> APPLICANT: Housman, David E.  
;> APPLICANT: Shaw, Duncan J.  
;> APPLICANT: Harley, Helen G.



APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,699A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/422,706  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5840A2  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-422-699A-12

Query Match 0.7%; Score 18; DB 2; Length 2726;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccgccgccc 263  
|||||  
DB 2173 GCCCTCCCTCCCGGCC 2156

RESULT 13  
US-08-422-706B-12/c  
Sequence 12, Application US/08422706B  
Patent No. 5977333  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.

TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,706B  
FILING DATE: 14-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/284,543  
FILING DATE: 08-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-422-706B-12

Query Match 0.7%; Score 18; DB 2; Length 2726;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctcccgccgccc 263  
|||||  
DB 2173 GCCCTCCCTCCCGGCC 2156

RESULT 14  
US-08-484-044-11/c  
Sequence 11, Application US/08484044  
Patent No. 555282  
GENERAL INFORMATION:  
APPLICANT: Caskey, C. T.  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Friedman, David L.  
APPLICANT: Pizzuti, Antonio  
APPLICANT: Fenwick, Raymond G.  
TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy

```

;
; APPLICATION NUMBER: US/08/422.699A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422.706
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/023.612
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/839.255
; FILING DATE: 20-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/01545
; FILING DATE: 19-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00253
; FILING DATE: 05-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9202485.0
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5830A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_difference
; LOCATION: replace(518..3323, "")
; OTHER INFORMATION: /standard_name= "CDNA 41"
;
; NAME/KEY: misc_difference
; LOCATION: replace(769..3323, "")
; OTHER INFORMATION: /standard_name= "CDNA 28"
;
; US-08-422-699A-10

```

```

Query Match 0.7%; Score 18; DB 2; Length 3323;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctccctcccgcc 263
| | | | | | | | | | | | | | | |
Db 2732 GCCCTCCCTCCCTCCCGGCC 2715

```

Search completed: July 23, 2002, 19:57:14  
Job time: 5709 sec

```

;
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 1301 McKinney, Suite 5100
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77010-3095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484.044
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/019.940
; FILING DATE: 19-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul, Thomas D.
; REGISTRATION NUMBER: 32,714
; REFERENCE/DOCKET NUMBER: D-5443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713/651-5325
; TELEFAX: 713/651-5246
; TELEX: 762829
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Gouble
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-484-044-11

```

```

Query Match 0.7%; Score 18; DB 1; Length 3182;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 gccctccctccctcccgcc 263
| | | | | | | | | | | | | | | |
Db 2637 GCCCTCCCTCCCTCCCGGCC 2620

```

```

RESULT 15
US-08-422-699A-10/c
; Sequence 10, Application US/08422699A
; Patent No. 5955265
; GENERAL INFORMATION:
; APPLICANT: Brook, J. David
; APPLICANT: Housman, David E.
; APPLICANT: Shaw, Duncan J.
; APPLICANT: Harley, Helen G.
; APPLICANT: Johnson, Keith J.
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC
; TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02713
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:26:55 ; Search time 2033.33 seconds  
(without alignments)  
17484.117 Million cell updates/sec

Title: US-10-041-770-1  
Perfect score: 2634  
Sequence: 1.atggagaactggactggcag.....cctcagcctttccagcatag 2634

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estim:\*
  - 5: em\_estor:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_htc:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_htc:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	360	13.7	437	10	BM148104
2	328	12.5	370	10	EG005397 PMI-GN018
3	295	11.2	891	9	AL513944
C 4	287	10.9	360	10	BF993048
5	286	10.9	307	10	BF989747
C 6	282	10.7	309	9	AI637480
7	255	10.1	324	10	BF993027
8	248	9.4	373	10	BF987707
9	246	9.3	507	10	BF986444
10	222	8.4	625	10	GC438243
C 11	211	8.0	262	10	BF989357
12	205	7.8	711	10	B1836313
C 13	193	7.3	193	10	BF987087
14	188	7.1	286	10	D78761
15	168	6.4	183	10	BF997839
16	167	6.3	214	10	BF997840
17	142	5.4	216	10	BF367876

18 110 4.2 449 10 BG900891  
19 107 4.1 865 10 BI094238  
20 103 3.9 375 10 EG008790  
21 94 3.6 327 10 BF993841  
22 87 3.3 389 10 BG008820  
23 86 3.3 266 10 BM150365  
24 86 3.3 273 10 BF997426  
25 83 3.2 419 10 D63308  
26 65 2.5 297 10 BF363794  
27 52 2.0 274 10 BM148104  
28 41 1.6 123 10 BF994270  
29 39 1.5 125 10 BG001393  
30 39 1.5 495 9 AW230557  
31 39 1.5 513 9 AW763191  
32 38 1.4 127 10 BI012353  
33 38 1.4 390 9 BF368874  
34 34 1.3 597 9 AI683106  
35 32 1.2 592 10 BM391602  
36 32 1.2 866 10 BI330332  
37 30 1.1 517 10 BE808908  
38 29 1.1 480 10 BF651524  
39 29 1.1 571 10 BE757541  
40 28 1.1 188 10 BG242515  
41 26 1.0 360 9 AW482344  
42 26 1.0 379 9 BB857367  
43 26 1.0 446 9 BB862174  
44 26 1.0 483 9 BB856356  
45 26 1.0 498 9 BB855583

ALIGNMENTS

RESULT 1  
BM148104  
LOCUS  
DEFINITION

BM148104 437 bp mRNA linear EST 30-NOV-2001  
TCAAP1Q8352 Pediatric acute myelogenous leukemia cell (FAB M1)  
Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP8352, mRNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM  
human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Contact: Dr. Judith F. Margolin  
Texas Children's Cancer Center and Human Genome Sequencing Center  
at Baylor College of Medicine  
1102 Bates, MC3-3320 Houston, TX 77030, USA  
Tel: 832-824-4536  
Fax: 832-825-4038  
Email: clones@txccc.org  
Seq primer: M13 primer.  
Location/Qualifiers

1. 437  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TCAAP8352"  
M1) Baylor-HGSC project-TCAA"  
/sex="male"  
/tissue\_type="leukopheresis"  
/cell\_type="myeloid cell"  
/dev\_stage="pediatric 6 years"  
/lab\_host="DH10B"  
/note="vector: lambda pSB; Site\_1: BamHI; Site\_2: EcoRI;  
First strand cDNA was primed with an anchored

FEATURES  
source

BM148104 TCAAP1Q83  
BF994270 CM2-GN016  
BG001393 MR3-GN018  
AW230557 uc6sall.y  
AW763191 ur60h06.y  
BI012353 PMI-ENO06  
BF368874 RC6-GN007  
AI683106 tx01d11.x  
BM391602 UI-R-DY0-  
BI330332 602883024  
BE808908 214215 MA  
BF651524 274324 MA  
BE757541 212006 MA  
BG242515 602354320  
AW482344 42347 MAR  
BB857367 BB857367  
BB862174 BB862174  
BB856356 BB856356  
BB855583 BB855583

XhoI-oligo(dN) primer [5'GGAGAGACTCGAGCGCGCGAGAGAG(T)VN  
3': V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand  
was primed with a BamHI-dC primer  
[5'AGAGAGTCGATCGCGCGCGCAATAATAAT(C) 3'].  
Double-stranded cDNA was then digested with BamHI and XhoI  
and directionally cloned into the BamHI and SalI sites of  
lambda PSB vector. Library went through one round of  
normalization. Library was constructed by Wei Yu at RIKEN  
of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,  
Itoh M, Nagaoka S, Sasaki, Okazaki Y, Muramatsu M,  
Schneider C, Hayashizaki Y, High efficiency selection of  
full-length cDNA by improved biotinylated cap trapper.,  
DNA Res 4: 1, 61-6, Feb 28, 1997).  
DNA Res 4: 1, 61-6, Feb 28, 1997).  
76 a 151 c 131 g 79 t  
BASE COUNT

Query Match	13.7%	Score 360	DB 10	Length 437
Best Local Similarity	100.0%	Prod. No. 4.6e-161		
Matches 360	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	17	gcaggccctgctgtatctgctgtctgtctgtctctccctccagctctgcttgatcagg	76	
Dd	14	GCAGCCCTGCTGTATCTGCTGTCTGTCTGTCTCTCCCTCAGCTGCTGTGATCAGG	73	

Qy	77	aggattatccggacactctttcagacacctacagaggagcgccaggccccaaagtg	136
Db	74	AGGTGTTTCGGACACTCTCTTACAGACCCTACAGAGAGGCCAGCGCCCCGAAGGTG	133
Qy	137	tctggggacccttgggtccagtgggcctcttgcctccagccctcggggtgggggtgcagc	196
Db	134	TCTGGGGAACCTTGGTCCAGTGGGCCTTCTGCTCCAGCCTGGGGGTGGGGTGCAGC	193
Qy	197	gcaggagcgcgacatgctcagtcctcctcagtcgagctccaccgccagtctgcctccctcccctc	256
Db	194	GCAGAGCGGGACATGTACGTCCCTCAGTGCAGTCCACCCGAGTCTGCCCTCCCTC	253
Qy	257	ccggcgcgcccaagacatcacagaacctctctccccggggcgagggtccacagaccccaga	316
Db	254	CCGGCGCCCCAAGACATCCAGAAGCCCCTCTCCCCGGGGCCAGGGTCCCAAGACCCCA	313
Qy	317	cttctccagaaaccttccccttgttacaggacacagctctcggggaagggtggccacttc	376
Db	314	CTTCTCCAGAAACCTTCCCCTTGTACAGGACACAGCTCTCGGGGAAGGGGTGCCACTTC	373

RESULT	2	370 bp	mRNA	linear	EST 24-JAN-2001
LOCUS	BG005397				
DEFINITION	PM1-GNC180-241100-003-e05	GN0180	Homo sapiens	CDNA	
ACCESSION	BG005397				
VERSION	BG005397.1				
KEYWORDS	EST.	GI:12447519			
SOURCE	human.				

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 370)

REFERENCE  
Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsumura,A., Bata,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL  
MEDLINE  
COMMENT

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml3.pl?tl=PMI&t3=PMI-GN0180-241100-003-e05&t3=2000-11-24&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 370.  
Location/Qualifiers  
1. 370  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT

80 a	127 c	98 g	64 t	1 others
------	-------	------	------	----------

performed under low stringency conditions." Reverse transcription of tissue mRNA and cDNA amplification were Research) profiles into the pUC 18 vector. Reverse application No. 196,776 - Ludwig Institute for Cancer products derived from ORESTES pCK (O.S. Letters Patent / Site:2: SmaI; A mini-library was made by cloning

[illegible]

RESULT	3	
AL513944		
LOCUS	AL513944	891 bp mRNA linear EST 13-FEB-2001
DEFINITION	AL513944 LTI_NFL006.PL2 Homo sapiens cDNA clone CLOBA010ZE04 5 prime mRNA sequence.	
ACCESSION	AL513944	
VERSION	AL513944.1	GI:12777438
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1. (Bases 1 to 891)	
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.	
TITLE	Full-length cDNA libraries and normalization	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Genoscope	

```

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBAR010ZE04"
/clone_lib="LTI.NFL006.PL2"
/tissue_type="Placenta"
/notes=Vector: PCMVSPORT 6; Site1: NotI; 1st strand cDNA
was primed with a NotI-oligo(drr) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT      162 a   305 c   265 g   155 t   4 others
ORIGIN
nucp:nuclengch:inv:trayen:com

Query Match      11.2%; Score 295; DB 9; Length 891;
Best Local Similarity 99.4%; Pred. No. 6.4e-130;
Matches 515; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 atggagaactggaactggcaggccctggctctatctctgctctgtctctgtctctcctcctcag 60
    |||
Db 222 ATGGAGAACTGGACTGGCAGGCCCTGGCTGTATCTGCTGCTGCTTCTGTCCCTCCCTCAG 281
    |||

QY 61 ctctgcttggatcagagaggtgtgttcgggacactctcttcagacacctacagagagggc 120
    |||
Db 282 CTTCTGCTTGGATCAGGAGGTGTTTTCGCGGACACTCTCTTCAGACACCTACAGAGGAGGC 341
    |||

QY 121 cagggcccccgaagtgtctgggaaccttgggtccagtggcctcttgtccagacctgc 180
    |||
Db 342 CAGGGCCCCGAA - GGTCTGGGGACCTTGGGTCCAGTGGGCTCTTGTCTCCAGCCCTTGC 400
    |||

QY 181 ggggtgggggtgcagcgagagccggacatgtcagctccctacagtgcagctccacccg 240
    |||
Db 401 GGGGTGGGGGTGCAGCGCAGGAGCGGACATGTCACTCCCTACAGTGCAGTCCACCCG 460
    |||

QY 241 agtctgcacctctctccggcccccgaagacatccagaagcctctctcccgggggcag 300
    |||
Db 461 AGTCTTCHCTTCCCTCCCGGCCCCCAAGACATCCAGAGCCCTCTCTCCCGGGGGCAG 520
    |||

QY 301 ggtccagaccocgaactcttcagaaacctccctctgtacagagacacagtcctcgggga 360
    |||
Db 521 GGTCCAGACCCCGACACTTCTCCAGAAACCCCTCCCTTTGACAGGACACAGTCTCGGGGA 580
    |||

QY 361 aggggtggccacttcgaggtcccgcttcccacctaggagagagagaccagagatt 420
    |||
Db 581 AGGGGTGGCCCACTTCGAGGTCCCGCTTCCCACTAGGGAGAGAGGAGACCCAGGAGATT 640
    |||

QY 421 cgagcgggcagaggtccggcttcgagaccocatcaagccaggaaatgtctcggttatgg 480
    |||
Db 641 CGAGCGGGCAGAGGTCCCGGCTTTCGAGACCCCATCAAGCCAGGAATGTTCGGTTATGGG 700
    |||

QY 481 agagtgcctttgtcattgcceactgcacgggaacgcag 518
    |||
Db 701 AGAGTGCCTTTGTGATTTGCCACTCAACCGGAACCGGAG 738
    |||

```

[illegible]

ORGANISM	
REFERENCE	
AUTHORS	

*Homo sapiens*  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 360)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Negai, M.A., da Silva, W. Jr., Zagzo, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matucko, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P., Buchner, P., Jorgensen, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**TITLE**

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**JOURNAL**

JOURNAL  
MEDLINE  
COMMENT  
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.

## COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil.

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimmons@ludwig.org.br](mailto:asimmons@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?fl=IL5&c2=IL5-GN0178-311000-204-fl0&t3=2000-10-31&t4=1>)  
Seg primer: puc 18 forward  
High quality sequence stop: 305.

## FEATURES

```

FEATURES
  source
    Location/Qualifiers
      1. 360
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="GN0178"
        /dev_stage="Adult"
        /note="Organ: Placenta_normal; Vector: puc18; Site_1: SmaI
          ; Site_2: SmaI; A mini-library was made by cloning
          products derived from ORESTES PCR (U.S. Letters Patent
          application No. 196,716 - Ludwig Institute for Cancer
          Research) profiles into the pUC 18 vector. Reverse
          transcription of tissue mRNA and cDNA amplification were
          performed under low stringency conditions."
      59 a      80 c      142 g      78 t      1 others
BASE COUNT
ORIGIN

```

25

BASE COUNT		ORIGIN	
59 a	80 c	142 g	78 t
		1 others	

Query Match 10.9%; Score 287; DB 10; Length 360;  
Best Local Similarity 99.7%; Pred. No. 3.9e-126;  
Matches 337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 492 tgcattgcactgcaccggaacgcgcaggacacctcgagccccaccagatcttgagctgtc 551

Db 338 TGCATTGCCACTGCACCGGAACCGCAGGCACCCCTCGGAGCCCCACCCAGATCTGAGCTGTC 279

Qy 552 cctgatctctcttagaggggaagaggctattccgtccctactccaagagcagagccatt 611

Db 278 CCTGATCTCTCTAGAGGGGAGAGCCCTATTCGGTCCCTACTCCAGAGCAGAGCCATT 219

Qy 612 ctccgaaacggcgagcccccaactgagctcctccacagaactgtgtgtccaccccc 671

Db 218 CTCCGGAAACGGCAGCCCCCAAACTGAGTCCCTCCACAGAACTGtGTGTCCACCCCC 159

Qy 672 atcccccaagcagaacctctatagccctgaactgctcagacagagtggtgccccagaac 731

Db 158 ATCCCCCAAGCAGAACCTCTTAAGCCCTGAAACTGCTCAGACAGAGGTGGCCCCCAGAAC 99

Qy 732 caggcctgccccctacggcatcacccagagcccaggcctctggcacagagccccctc 791

Db 98 CAGGCGTCCCCCTACGGCATACCCCCAGAGCCAGGCCCTCTGGCACAGAGCCCCCTC 39

Qy 792 accaagcactccttaggagaaaggtggcctttccgtg 829

Db  
38 ACCCAGCACTCCTAGGAGAGGTGGCTTCTTCCGTG 1

```

RESULT 5
BF989747 307 bp mRNA linear EST 23-JAN-2001
LOCUS ILS-GN0178-231000-195-f05 GN0178 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF989747
ACCESSION BF989747.1 GI:12396072
VERSION EST
KEYWORDS
SOURCE human.
ORIGIN Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 307)
Dias Neto,E., Garcia Correa,R., Varjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.D. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL5&t2=IL5-GN0178-
231000-195-f05&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 289.
FEATURES
Location/Qualifiers
1..307
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0178"
/dev_stage="Adult"
/Note="Organ: Placenta_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 64 a 106 c 87 g 50 t
ORIGIN
Query Match 10.9%; Score 286; DB 10; Length 307;
Best Local Similarity 100.0%; Pred. No. 1.2e-125; Indels 0; Gaps 0;
Matches 286; Conservative 0; Mismatches 0;
QY 363 gggtgcccacttcaggtccgcgttcccactaggagagagagaccaggagattcg 422
Db 1 GGTGTGCCCACTTCAGAGTCCCGCTCCACCTAGGAGAGAGAGACCAGAGATTCCG 60
QY 423 agcgccaggaaggtcccggttcgagagcccatcaagccaggaagtgtcggttatggag 482
Db 61 AGCGGCCAGGAGGTCCCGCTTCAGAGACCCCATCAAGCCAGGAAATGTCGGTATGGGAG 120
QY 483 agtgccttttgattgccattgcacgcgaacccgagggcaccttcgagccaccagcagtc 542
Db 121 AGTGCCTTTGCAATTGCCATGCACCGGAACCGGACCCCTCGAGAGCCACCCAGATC 180
QY 543 tgagctgtccctgatctcttcttagaggggaagaggtattccgtccctactccaagac 602
Db 181 TGAGCTGTCCCTGATCTCTTCTAGAGGGAAGAGGCTATTCCGTCCCTTACTCCAAAGAC 240

```

Qy	2533	tggtccacagcagctggagctcaaggtgagcccggaaccccccagccatctctcattc 2592
Db	129	TGGTTCCACAGCGACTGAGCTCAAGGTGAGCCCGGAACCCCGAGCATATCCTGCATC 70
Qy	2593	ctgggttaaccatgcccagggacacctcagctcttccagcatag 2634
Db	69	CTGGGTACCAATGCCAGGACACCTCAGCCTTCCASCATAG 28
RESULT	7	
LOCUS	BF993027	
DEFINITION	IL5-GN0178-311000-204-e09 GN0178 Homo sapiens cDNA, mRNA sequence.	EST 23-JAN-2001
ACCESSION	BF993027	
VERSION	BF993027.1	GI:12399350
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 324)	
DIAS NETO, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.J. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5st2-IL5-GN0178-311000-204-e09st3-2000-10-31&t4=1) Seq primer: puc 18 forward High quality sequence stop: 315. Location/Qualifiers	
FEATURES	source	
BASE COUNT	69 a 115 c 85 g 55 t	
ORIGIN		
Query Match	10.1%; Score 265; DB 10; Length 324;	
Best Local Similarity	99.7%; Pred. No. 1.3e-115;	
Matches	315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	363	gggtgcccacttcaggtccctccacctcagggagagagagagagagagagattcgc 422
Db	1	GGGTGGCCCACTTCGAGGTCCCGCTCCACCTPAGGAGAGAGAGAGAGGATTCG 60
Qy	423	agcgccagagggtcccggttcgagaccccatcaagccaggaatgttcggtatggag 482
Db	51	AGCGGTCAGAGGTCGCTTCGAGACCCCATCAGCCAGGAATGCTCGGTATGGAG 120

[illegible]

BASE COUNT	83 a	158 c	171 g	95 t	performed under low stringency conditions."	
ORIGIN						
Query Match	9.3%; Score 246; DB 10; Length 507;					
Best Local Similarity	100.0%; Pred. No. 1.7e-106; Mismatches 0; Indels 0; Gaps 0;					
Matches 246; Conservative						
QY	1	atggagaactggactggcagccctggctgtatctctgtctgtctctctccctccctcag	60			
DB	232	ATGAGAACTGGACTGGCAGGCCCTGGCTGTATCTGCTGTCTTCTGTCCCTCCCTCAG	291			
QY	61	ctctgttgatcagaggtgtgttcggacactctcttcagacactacagaggggc	120			
DB	292	CTCTGCTTGGATCAGGAGGTGTGTTCGGACACTCTCTTCAGACACTACAGAGGAGGC	351			
QY	121	caggggcccaagtgctctgggacacttgggtgcctgggtgcctctgtctccagccctgc	180			
DB	352	CAGGCCCCGAAGTGCTCTGGGACCTTGGTCCAGTGGGCTCTTGTCTCCAGCCCTGC	411			
QY	181	gggtgggggtgcagcgcagggacgcggacatgtcagctccctacagtgcagctccaccg	240			
DB	412	GGGTGGGGTGCAGCCAGGAGCCGGACATGTACGTCCTCCTACATGAGCTCCACCCG	471			
QY	241	agtcgtg	246			
DB	472	AGTCGTG	477			
RESULT 10						
BG438243						
LOCUS	602490152F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4622245 5',					
DEFINITION	625 bp mRNA linear EST 14-MAR-2001					
ACCESSION	BG438243					
VERSION	BG438243.1 GI:133444749					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 625)					
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .					
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rcapbs-r@mail.nih.gov Tissue Procurement: DCTD/DTP/Gazdar cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCMI386 row: g column: 14 High quality sequence stop: 406. Location/Qualifiers 1. .625 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4622245" /clone.lib="NIH_MGC.18" /tissue_type="large cell carcinoma" /lab_host="DH10B (phage-resistant)" /note="Organ: lung; Vector: pORF7; Site: 1; XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."					
FEATURES	source					



```

BASE COUNT      188 a      161 c      174 g      102 t
ORIGIN

Query Match      8.4%; Score 222; DB 10; Length 625;
Best Local Similarity 100.0%; Pred. No. 5.4e-95;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1464 ggggaacctactgacgagggcccccctggcgtatcagaagatcttgtgattccagc 1523
Db 51 GGGGAACCTCAGTACCGAGGGGGCCCCCGGCTTTCAGAAATCTTGTGGATTCCAGC 110

QY 1524 gggagccttggcgtccagattgccagctccggcctagctccaaactactggtcacttgg 1583
Db 111 GGGAGCCTTCGGCTCCAGATTGCCAGCTCCGGCTAGTCCCAACTACCTGSCACTTCG 170

QY 1584 tggccctggggcggtccatcaatgaatggaaactgggctgtgattcccccgtgctcta 1643
Db 171 TGGCCCTGGGGCGGGTCCATCATCAATGGGAACTGGGCTGTGGATCCCGCTGGTCCCTA 230

QY 1644 caggccgcgggagccgtcttctgatataacgctccctccag 1685
Db 231 CAGGCCCGGGGACCGCTCTTCGATATACCGCTCCCTCCAG 272

RESULT 11
LOCUS      BF989357/c      262 bp      mRNA      linear      EST 23-JAN-2001
DEFINITION      IL5-GN0179-201000-185-e11 GN0179 Homo sapiens cDNA, mRNA sequence.
ACCESSION      BF989357
VERSION      BF989357.1 GI:12395682
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 262)
AUTHORS      Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE      20202663
COMMENT      Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               Project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&st=IL5-GN0179-
               201000-185-e11&t3=2000-10-20&t4=1)
               Seq primer: puc 18 forward
               High quality sequence stop: 262.
               Location/Qualifiers
                   1..262
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone_lib="GN0179"
                       /dev_stage="Adult"
                       /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                               ; Site_2: SmaI; A mini-library was made by cloning
                               products derived from ORESTES PCR (U.S. Letters Patent
                               application No. 196,716 - Ludwig Institute for Cancer
                               Research) profiles into the pUC 18 vector. Reverse
                               transcription of tissue mRNA and cDNA amplification were

```

```

BASE COUNT      56 a      86 c      79 g      41 t
ORIGIN

Query Match      8.0%; Score 211; DB 10; Length 262;
Best Local Similarity 99.6%; Pred. No. 9.1e-90;
Matches 261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1432 gggggtgatgttacctgtgccttgttcggggaaacctactgacgagggggcccc 1491
Db 262 GGGGGTGTGATGATCTACCTGTGCCTTGTTCGGGGAACTCCTACCGAGGGGGCCCC 203

QY 1492 ctgggctatcagaagatcttggattccagcgggagccttgcggctccagattgccag 1551
Db 202 CTGGGCTATCAGAAGATCTTGTGGATTCCAGCGGGAGCGCTTCGGCTCCAGATTCCCGAG 143

QY 1552 ctccggcctagctcaaacctacctgaccttgccttgccttggggcggtccatcaat 1611
Db 142 CTCGGCCCTAGCTCTCACTACCTTGCACATTCGTGCCCTTGGGGCGGGTCCATCATCAT 83

QY 1612 gggaaactgggctgtggatccccctgggtcctacagggcgcgggagaccgttttgatat 1671
Db 82 GGGAACTGGGCTGTGGATCCCGTGGGTCTACAGGCGCGGGACCGTCTTTCGATAT 23

QY 1672 aacctctctccagggaggagg 1693
Db 22 AACCGTCTCCCGAGGAGGAGG 1

RESULT 12
LOCUS      BI836313/c      711 bp      mRNA      linear      EST 04-OCT-2001
DEFINITION      603083995F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5223358 5',
               mRNA sequence.
ACCESSION      BI836313
VERSION      BI836313.1 GI:15947863
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 711)
AUTHORS      NIH-MGC http://mgi.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-remail.nih.gov
               Tissue Procurement: Life Technologies, Inc.
               CDNA Library Preparation: Life Technologies, Inc.
               CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM11561 row: m column: 23
               High quality sequence stop: 711.
               Location/Qualifiers
                   1..711
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone="IMAGE:5223358"
                       /clone_lib="NIH_MGC_120"
                       /lab_host="DH10B"
                       /note="Organ: pooled pancreas and spleen; Vector:
                               pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                               source anonymous pool of spleen and pancreas from 28 yo
                               male. Library is oligo-dT primed and directionally cloned
                               (EcoRV site is destroyed upon cloning). Average insert
                               size 1.5 kb, insert size range 1-2.5 kb. Library is
                               normalized and enriched for full-length clones and was
                               constructed by C. Gruber (Invitrogen). Research Genetics
                               tracking code 025. Note: this is a NIH_MGC Library."

```

BASE COUNT	151 a	233 c	222 g	104 t	1 others
Query Match	7.8%; Score 205; DB 10; Length 711;				
Best Local Similarity	100.0%; Pred No. 7.5e-87;				
Matches 205; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	2178	ctggagagctggcagtgagacatctctgcagccgctctctgtgcccgcgcaccagcacgcg	2237		
Db	210	CTGGAGAGCTGGCAGTGGACATCTCTCAGCCGCTCTGTGGCCGCGCACCCAGCACCG	151		
QY	2238	ccagctcagtcgcgcgcaggaatttgggggggtgactctctcgtgtgcccccagagcgtg	2297		
Db	150	CCAGCTGCAGTCCCGCAGGAATTGGGGGGGTGGCTCTCGGTGCCCCCGAGGGCTG	91		
QY	2298	tggacatctccccggcccaacatccaccagcttctgcagctgcgcctctgtggccattg	2357		
Db	90	TGGACATCTCCCGCGGCCAACATCACCAGTCTTGCCAGTGGCGCTCTGTGGCCATTG	31		
QY	2358	ggaagtggctctccttagaccag	2382		
Db	30	GGAAGTGGCTCTCTCTGGAGCCAG	6		
RESULT 13					
BF870887					
LOCUS	BF870887 193 bp mRNA linear EST 17-JAN-2001				
DEFINITION	IL0-ET0152-301000-501-c03 ET0152 Homo sapiens cDNA, mRNA sequence.				
ACCESSION	BF870887				
VERSION	BF870887.1 GI:12261017				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 193)				
	Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL04t2-IL0-ET0152-301000-501-c03&t3=2000-10-30&t4=1) Seq primer: puc 18 forward High quality sequence stop: 193. Location/Qualifiers 1..193 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="ET0152" /dev_stage="Adult" /note="Organ: lung tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under				
FEATURES					
source					

Db 76 CATTGGGAAGTTGGCTCTCCTTTGGAGCCAGTGTCCCTGGGTCGGGCGGCCGAGAGA 135  
Qy 2413 agccggcaggttcgctgtgttgggaacacacggtgatgaagtgcgag 2460  
Db 136 AGCCGGCAGGTTGCTGTGTGGGAACAACGGTGATGAAGTGAGCGAG 183

Search completed: July 23, 2002, 19:00:50  
Job time: 5635 sec

Qy 2407 cagagaagccgaggttcgctgtgttgggaacacacggtgatgaagtgcgagcagag 2466  
Db 121 CAGAGAAGCCGCGAGTTCGCTGTGTGGGAACAACGGTGATGAAGTGAGCGAGCAGAG 180  
Qy 2467 tctgcgtc 2474  
Db 181 TGTGGGTC 188

RESULT 15  
BF997839  
LOCUS BF997839 183 bp mRNA linear EST 23-JAN-2001  
DEFINITION MR2-GN0127-091100-005-402 GN0127 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF997839  
VERSION BF997839.1 GI:12404162  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 183)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=MR2&t2=MR2-GN0127-  
091100-005-d02&t3=2000-11-09&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 183.  
FEATURES  
Location/Qualifiers  
1..183  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0127"  
/dev\_stage="Adult"  
/note="Organ: placenta normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 28 a 56 c 63 g 35 t  
ORIGIN

Query Match 6.4%; Score 168; DB 10; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3.4e-69;  
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2293 cgcgtggagacatccccccgcccacacatccagcttcgagctgcctctgtgac 2352  
Db 16 CGCTGTGGACATCCCGCCGCCCAACATCACCGATCTTGGCAGCTGCGCCCTCTGTGGC 75

Qy 2353 cattgggaagtggctctctcttgagccagtgctcccgcggtgcgcccgggcccagaga 2412



TRE

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 179:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-179

Query Match 1.6%; Score 43.2; DB 4; Length 400;
Best Local Similarity 49.1%; Pred. No. 0.14;
Matches 114; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 2205 cagccgctctgtgcccgcgcacccagcagccagctgagcttgacatctcccccgcccaacatcac 2264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 CGCGCGGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227
QY 2265 ggggggtgctctcggtgccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 CGCGCGCGGTTCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 167
QY 2325 ccagcttgcagctgcccgcctctgtggccattgggaagtggctctctctggagccaagt 2384
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 CGCGCGTCACCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 107
QY 2385 ctccgtgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2436
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 GCGCCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 55

RESULT 5
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU

```

[illegible]

```

; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 00303/015002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-076-089-1

Query Match 1.6%; Score 42.2; DB 1; Length 1443;
Best Local Similarity 44.4%; Pred. No. 0.39;
Matches 170; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 17 gcaggccctggctgtatctgctgctgttctgtcttcctcctcagctctgcttggaatcagg 76
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 412 GCAGCAGACACCTTGTAAACGCTCTGCTGAGTCGCTGCGCTGCCCTCAGCTGACGAGTCTT 353
QY 77 agtggtttccggagactctcttcagacactacagagagggccagggccccccaaggtg 136
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 352 CGAGTCTCTCGGGCAGTCCAGCTGGGACCTGGTCCCGTCCCGCGCGCGCGCGCG 293
QY 137 tctggggaccttgggtccagtgggcctctctgtcccaagccctgcggggtgagggcgc 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 292 TCAGGGCACCCGGGGTTCAGCGCCGCTGCAAGTCGCGCTCTCCACCGCATGCTGCGCG 233
QY 197 gcaggagcggagatgtcagctcctcacagtgcagctccaccgcaggtctgccccctccc 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 232 GGTGCAGCGGGGGGGGGCCCGAGGGTGTGCTGCCCGACGGGCTCGCTCTCTGGC 173
QY 257 ccggggcccccagaacatccagaagccctctctccccggggccaggggtccacagacccca 316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 172 CCCCACCGCGGCTCCCGCCGCTCCGTCGCGCGCGCTTCAGGGTCAATCGGGTCCGGGA 113
QY 317 ctctccagaaacctccctcttgacaggacagctctcggggaaggggtgcccacttc 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 112 CGGTCCACTCGCACACACACACGCTCAGGACTAGGATCCGGATTAGGAGGCCACGCATC 53
QY 377 gaggtccgcgttcccaactaggg 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 52 GTAGTCCGGAGCTCGGACTTTAGG 30

RESULT 7
US-08-707-200-1/c
; Sequence 1, Application US/08707200
; Patent No. 5891430
; GENERAL INFORMATION:
; APPLICANT: C. Ronald Kahn, M.D.
; APPLICANT: Christine Revnet, Ph.D.
; TITLE OF INVENTION: DIABETOGENE RAD: A TYPE II DIABETES
; TITLE OF INVENTION: SPECIFIC GENE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; Zip: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08707,200

```







QY 204 ccggacatgtcagtcctcctacagtcagtcacccagtgctgcccctccctcccccggcc 263  
 Db 236 ACCGACTTCCAGCCCGGGTCCCGCCACACCCGAGCGCGCGCCCTGCGCGCGAAC 295  
 QY 264 cccaagacatcca 276  
 Db 296 CCGTGGCTGCCA 308

RESULT 13  
 US-08-483-533-4  
 ; Sequence 4, Application US/08483533  
 ; Patent No. 6172047  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Roizman, Bernard  
 ; APPLICANT: Chou, Joany  
 ; TITLE OF INVENTION: Method for Treating Tumorigenic  
 ; TITLE OF INVENTION: Diseases  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION NUMBER: US/08/483,533  
 ; FILING DATE: 07-MAR-95  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION NUMBER: 08/419,853  
 ; FILING DATE: 11-APR-95  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/861,233  
 ; FILING DATE: 31-MAR-92  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Zeller, James P.  
 ; REGISTRATION NUMBER: 28,491  
 ; REFERENCE/DOCKET NUMBER: 28097/32742  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 595 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-483-533-4

Query Match 1.5%; Score 39.4; DB 4; Length 595;  
 Best Local Similarity 48.4%; Pred. No. 1.4;  
 Matches 109; Conservative 0; Mismatches 116; Indels 0; Gaps 0;  
 QY 162 ctcttgtccacgacctgcggggtgcagcgagagccgagacatgtcagctccc 221  
 Db 122 CGCTCGACGCGGGGAGGGGCGCGAGACCCCGGACCCCGCGACCC 181  
 QY 222 tacagtgcagtcacccgagctgtgcccctccctcccgcccccaagacatccagaagc 281  
 Db 182 GCGACCCCGCGACCCCGCGACCCCGCGACCCCGCGACCCCGCGACCC 241  
 QY 282 cctctctcccccggggccagggtccacagacccagacctctctcagaagaacccctcccttgta 341

Db 242 CCCCGAGCCCGCGGGTGGCTTCTCCGCCACGCTCCGGGTGCGCACCTGGTGGTC 301  
 QY 342 cagacacagtcctcggggaagggtgagccacttcaggtccagc 386  
 Db 302 TGGCCTCGCCCGCCGCTGGCGCGCGGCTGTGGGCGCCGC 346

RESULT 14  
 US-09-283-471A-4  
 ; Sequence 4, Application US/09283471A  
 ; Patent No. 6340673  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Roizman, Bernard  
 ; APPLICANT: Chou, Joany  
 ; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION NUMBER: US/09/283,471A  
 ; FILING DATE: 04-APR-1999  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION NUMBER: 07/861,233  
 ; FILING DATE: 31-MAR-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/419,853  
 ; FILING DATE: 11-APR-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/483,533  
 ; FILING DATE: 07-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Zeller, James P.  
 ; REGISTRATION NUMBER: 28,491  
 ; REFERENCE/DOCKET NUMBER: 27373/32742A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 595 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-09-283-471A-4

Query Match 1.5%; Score 39.4; DB 4; Length 595;  
 Best Local Similarity 48.4%; Pred. No. 1.4;  
 Matches 109; Conservative 0; Mismatches 116; Indels 0; Gaps 0;  
 QY 162 ctcttgtccacgacctgcggggtgcagcgagagccgagacatgtcagctccc 221  
 Db 122 CGCTCGACGCGGGGAGGGGCGCGAGACCCCGGACCCCGCGACCC 181  
 QY 222 tacagtgcagtcacccgagctgtgcccctccctcccgcccccaagacatccagaagc 281  
 Db 182 GCGACCCCGCGACCCCGCGACCCCGCGACCCCGCGACCCCGCGACCC 241  
 QY 282 cctctctcccccggggccagggtccacagacccagacctctctcagaagaacccctcccttgta 341



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 20:17:51 ; Search time 24.4 seconds  
(without alignments)  
877.920 Million cell updates/sec

Title: US-10-041-770-2  
Perfect score: 4895  
Sequence: 1 MENWGRPWLYLLLSLPQ.....PPAISILGNHAQDTSAPFA 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	284	5.8	788	2	US-08-918-914-4
2	232.5	4.7	1064	1	US-08-642-255-62
3	224.5	4.6	960	4	US-09-219-849-5
4	224.5	4.6	1185	4	US-09-041-886-23
5	220	4.5	882	4	US-09-413-814-78
6	216.5	4.4	1341	3	US-08-963-825-18
7	216.5	4.4	1341	4	US-09-570-573-18
8	216.5	4.4	1341	4	US-09-548-608-18
9	213	4.4	1055	1	US-08-642-255-72
10	210.5	4.3	1057	3	US-08-931-820-1
11	210.5	4.3	1321	2	US-08-317-310A-64
12	209	4.3	408	1	US-07-609-716-65
13	209	4.3	408	4	US-08-475-411A-65
14	209	4.3	408	4	US-08-478-029A-65
15	205.5	4.2	1078	3	US-08-963-825-21
16	205.5	4.2	1078	4	US-09-570-573-21
17	205.5	4.2	1078	4	US-09-548-608-21
18	204.5	4.2	1060	3	US-08-931-820-3
19	204.5	4.2	1418	3	US-08-963-825-20
20	204.5	4.2	1418	4	US-09-010-999-1
21	204.5	4.2	1418	4	US-09-570-573-20
22	204.5	4.2	1418	4	US-09-548-608-20
23	203	4.1	960	4	US-09-219-849-6
24	202	4.1	552	4	US-09-219-849-7
25	201	4.1	720	4	US-09-219-849-4
26	201	4.1	777	1	US-08-642-255-53
27	200.5	4.1	633	1	US-08-642-255-73

Sequence 4, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 3, Appl  
Sequence 52, Appl  
Sequence 3, Appl  
Sequence 120, Appl  
Sequence 31, Appl  
Sequence 19, Appl  
Sequence 19, Appl  
Sequence 19, Appl  
Sequence 2, Appl  
Sequence 5, Appl  
Sequence 114, Appl  
Sequence 26, Appl  
Sequence 84, Appl  
Sequence 14, Appl  
Sequence 49, Appl

# ALIGNMENTS

RESULT 1  
US-08-918-914-4  
; Sequence 4, Application US/08918914  
; Patent No. 5876963  
; GENERAL INFORMATION:  
; APPLICANT: Mitchell, Peter  
; APPLICANT: Hutchinson, Nancy  
; APPLICANT: Lawton, Michael  
; APPLICANT: Magna, Holly  
; APPLICANT: Yocum, Sue  
; APPLICANT: Murry, Lynn E.  
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/918,914  
; FILING DATE: Filed Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0369  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 788 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1070094  
; US-08-918-914-4



QY	419	ELNCRPRGRFFYHRTKXVDQDTLQOPGAPD1CVAGRCLSPGCDGILSGRRPDCGVC	478
Db	347	-----PPGSR-----GDPGPPGAP-----GPAGPPGSRDGP	382
QY	479	GDDSTCLRSGNLTRG-----GPLCYQKILWPAGALRLQIAQRPSNVLALRGPG	533
Db	383	-----PPGSRGDGPPGAPGAPG-----PGS-----RGDPGPPGAPG	419
QY	534	SIINGNAVDPPGSYRAGOTVFRYNNRPPREBKGSLSAEGTTPQDVYIMFOENPGV	593
Db	420	PGSRGD--PGPPGAPGAPG-----PPGSRDGP	464
QY	594	FYQYVSSPPPLENTPEPPYPQLOPELTRVEPPLAPAPRARTPGTLQROVRIPQMPA	653
Db	465	PGAPGAPGPPSGRDGPPGAPGAPGPPGSRGDGPPGAPGPPGSRGD-----PG	517
QY	654	PPHPRTPGLSGPAAFYWRVGHSAACSGXGVWRPIFLCISRSGBELDERSCAAGARPPA	713
Db	518	PPGAHGAPGPKGAHPAGPKGAHGAP-----AGPK-----GAPGA	561
QY	714	SPBPCHGTPCPTYWEAGETWTSRSCGPTQHRQLCROEFGGGSSVPPBRCGH-----	768
Db	562	GPPGSRGDGPPGAPGAPGPPGSRG--DPGPP-----GAPGAPGPPGSRDGP	610
QY	769	LPRPNITQSCQLRCLGHWEVGSFWSQCSYRCGRGQRSRQVRGNNNGDEVSEQECASPP	828
Db	611	APG-----AGPP-----GSRGD--PGPPGAPGA	633
QY	829	QPPSREACDMGPCTTAWFHSDMSSKVSPEPP	859
Db	634	GGPSRG--DPGPPGAPGAPGPPGSRDGP	663

```

RESULT      3
US-09-219-849-5
US-09-219-849-5
Sequence 5, Application US/09219849
Patent No. 6150081
GENERAL INFORMATION:
APPLICANT: VAN HEERDE, GEORGE V.
APPLICANT: VAN RIJN, ALEXIS C.
APPLICANT: BOWSTRA, JAN B.
APPLICANT: DE WOLF, FREDERIK A.
APPLICANT: MOOBROEK, ANDREAS
APPLICANT: WERTEN, MARC W.T.
APPLICANT: WIND, RICHELE D.
APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
TITLE OF INVENTION: PREPARATION THEREOF
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 5
LENGTH: 960
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Illustrative
OTHER INFORMATION: amino acid sequence
US-09-219-849-5

```

	Query Match	4.68;	Score	224.5;	DB.4;	Length	960;
	Best Local Similarity	25.28;	Pred.	Mis.3.7e-07;			
	Matches	194;	Conservative	25;	No. matches	333;	Indels
							227; Gaps
Qy	36	PTBEGQGEVGMVQWVWASCSQPGCVGQRRSRTCQLPTVOLHPSLPLPPRRPHREAL	95				
Db	95	PCSRDGPFGAPG-----AGPPE-----SRDGPFGAP-GPAGPFGSRDGPFGAP	140				
Qy	96	LPRG-----QGPRQTSPTLTYTSRGRGP-LRGFASHLGREETQEIARRSRRLD	150				

Db	141	GPAGPGSRDPGGPPGAG--PAGPPGSRDPGGPAGPAGPGS-----RD 184
Qy	151	PIKMGEGYGRVPFALPLHNRHRRSPRSELSLISSRGEAAIPSTTPRAEPPSAN-GS 209
Db	185	PGPPGAG-----PAGPP-----GSRDPGGPAGPAGPGPSRDPG 221
Qy	210	PQTELPTELSVHTPS-POAE--PLSPETAQTEVAPRTAPALRHHPAQAASGTEPPST 266
Db	222	PGAHPAGPAGKAHGPAGPAGKAHGPAGKGAHGPAGKAGP-----AGPGSRDPGPP 275
Qy	267	HSLGEGEFFRASPPRRPSQGWASPVAGRRDPFPSPVRGQOQG---GPMGTGGTP 323
Db	276	GAPGAG-PPGSRDPGPPGAPGAGPP-GSRDPGP-PGAPGAPGPGSRDPGPPGAGP- 331
Qy	324	HGPRLEDPQHBGAWLPLLSNGPHASSLSLFAFASP-IPRCSESEQLRACSOACPPE 382
Db	332	AGPGSRDPGPPGAGPA---GPPGSR----DFGPPGAPGAPPG-----SRDGPGP 378
Qy	383	QPDPRALQCAAFNSQEFMGOLYQWEPTEYQGSORCELCNCRPGFFYVYRHTEKVOGDTL 442
Db	379	APGPAG-----PPGSRDPGPPGAPGAPGAPGPGSR-----DPG--- 409
Qy	443	CQCAPDICVAGRCCLSPGCDIILG---SGRR-----PDGGCVGGDDSTCLRVSGNL 491
Db	410	-PPGAGPAGPPGSRDPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 468
Qy	492	TDRG--GPLYQYKILWTPAGALRLQIAQLRSPSNYLALRPGGSRSTINGNWAIVDPGSYR 549
Db	469	GPKGAHPAG-----PKGA-----HGPAKPKAHG-----PAGPKG 499
Qy	550	AGGIVFRYNRPREGKGESLSAEGPTTQPVDTYMTFQENBPVFYQYVISSPPPILENP 609
Db	500	APGPA---GPPGSRDPGPP-GAPGAPGP-----GSRDPGPPGAPGAPGPPGSRDPG 547
Qy	610	TPEFPVQLQPEILRVBPFLAP-----APRPARTPGLTQROVRIQOMPAP 654
Db	548	PPGAPGAPGPGSRDPGPPGAPGAPGPPGSRDPGPPGAPGAPGPPG-----RDPGPPGA 602
Qy	655	PHRPTPLGSPAAYWKRVGHSACSCKGVWRPFLICISRESGEELDRSCAAGAPPPAS 714
Db	603	PGPAGPPG-----RDPGPPGAPGAPGPPG-----SRDPG-----PGAPGPGAG 641
Qy	715	PEPCHGTPCPPIWEAGEWMTSCRSRCPGTQHRQLQCRQEFGGGSSVPP 763
Db	642	P-PGSRDPGPPGAPGAPGPGSRDPGPP-----GAPGAPGP 677

```

RESULT      4
US-09-041-886-23
Sequence 23, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: RabiZadeh, Sharroz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
City: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:

```

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: 8-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1185 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-041-886-23

Query Match 4.5%; Score 224.5; DB 4; Length 1185;  
Best Local Similarity 20.8%; Pred. No. 4.8e-07;  
Matches 212; Conservative 50; Mismatches 293; Indels 465; Gaps 52;

QY 16 LSLPOLCDEVLGSHSLQTEEGQEGYVWGWVWASQPCQGVQVRRST----- 70  
DB 96 LRPQSPDLSLGRSL-----NDGSG-----SDPRDIDQDNSTSPSIYS 137  
QY 71 -----COLTVQLHPSLPLPRP-----PRHPEA-----LLPRG----- 99  
DB 138 PGSVENDSSGLSQGPARYHPPLPPSPQPPDSTPROPEASFEPHPSVTGTGYHAP 197  
QY 100 -----QGP--RPQTSPTLPLRYTQSRGGGLGPGASHLGRERTQETIRAAARSRL 148  
DB 198 MEPTSRMFQAPPGAPPHPQLYP-----GGTGVLSGP----- 231  
QY 149 RDIPIKGMFYGRVFTALPHNRHRSPRSELSISRGGEAIPSPTPRAEPFSANG 208  
DB 232 --PMGPK--GGGAASSVGGPNGKQHP--PPTTPIIS--VSSGASGAPPTKPTTTPVGGN 284  
QY 209 SPQT-----ELPP-----TELSTVTPSQAE-----PL 231  
DB 285 LPSAPPANFPHVTNLPPPPALRNLNASAPPGGLGQALFGLHPLSPYANGMGGLP 344  
QY 232 SPETAQTEVAP-----RTRPAPLRHHPRAQASGTEP----- 262  
DB 345 GPEKGPT-LAPSPHSLPLPASSAPAPPMRFPYSSSSSSSSAAASSSSSSSSSASPPFASQ 403  
QY 263 --PSPHSLGEGGFRAPOPERR-----PS-----SQGWASQVAGR-----RPDPFP 303  
DB 404 ALPSYHSPPTPTSLSVSNQPKYQPSLPSQAVWSQFPPPPYPYGRLLANSNAHPGFP 463  
QY 304 -----SVPRGRGQQGQGWGTPGTHGPRLEPDP-----Q 333  
DB 464 PSTGAQSTAHPPVSTHHHHQOQQOQQOQQOHHGNSGPP-PPGAPPHLEGSSSHA 522  
QY 334 HPGAWLPILSN-----GPHASLSLFLAPSPIPRCSESLRACSOACPP 381  
DB 523 HPYAMSPSLGSLRPPYPPGPAHLPPHSHQVSYQAQPNG--PPVSSSSSSSSSTSGSYPC 580  
QY 382 EQDPDRALQCAAFNSQEFMQLYWEFTEVQGSOR-----CELNCRPRGFRFVYRHT 435  
DB 581 SHSP-----SQPGQAFYPPPPVPTVTTSSATLSVTIATVASSPAGYK----- 624  
QY 436 KVQDGLTQOPGAPDICVAGRCISLPGC-DGILSGRRPDCGCGGDDSTCLRVSNLTD 494  
DB 625 -----TASPPGPPPP--YGRAPSPGAYKATPPGYKP-----GSPPSFR----- 662  
QY 495 GGPLYGQKILWIPAGALRLQIAQLRPS-----NYIALRPGGGRSIIINGNAVDPPGYSR 549  
DB 663 GTPPYRGIS-PPAGP-----GTFKFGSPTVGVGPLPPAGPSGLPSL-----PP----- 705  
QY 550 AGTGVFRYNNRPREEGKESLSLNGTTPQDVYIMLFQENGVYQYVSSPPFLENP 609  
DB 706 -----PPAAPASGPPLSATQIKQEPABEY-----ETP-----ESPVPFPAKSP 742

RESULT 5  
US-09-413-814-78  
Sequence 78, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
TITLE OF INVENTION: heteropolyketide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 78  
LENGTH: 882  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-78

Query Match 4.5%; Score 220; DB 4; Length 882;  
Best Local Similarity 22.8%; Pred. No. 6.7e-07;  
Matches 165; Conservative 46; Mismatches 240; Indels 274; Gaps 35;

QY 29 SGHSLQTPTEEGQEGVW--GPMVQWASQSQ-----PCGVGVQVRRSRTCOLPTVQLHPSLP 83  
DB 281 STHASEPASLAEALGQLEAGHAVDWTARPRPPPARSSACRTRSSRGTSRRAAAP 340  
QY 84 LPP--RPPRHPALLPGQGPRTQTSPTETLYTQSGRGGPLRGA-----SHLGR 134  
DB 341 PPPSRRTPTFSR--PRTPSRNRSPARTPGPSRPRRPSARSAPAIWERPPRRDR 398  
QY 135 EETGEIARAARSRLDPTKPMFGYGRVPTALPLHRRHRPRSPRSELSSISRGEEAI 194  
DB 399 RPRLLRPARLARDP-----GDVPD--PRAARRAPAA----- 431  
QY 195 PSPTPRAEPFSANGSPQTELPPTLSVHTSPQAEPLSPETAQTEVAPTRPAPLRHHR 254  
DB 432 -GALPGADGRCARGGRRRAPRRGAAARVPPDRADP--PRRPAPVY-RAAPAVRRSAR 487  
QY 255 AQASGTEPSPSTHSLGEGGFRAFPQPRRPSQGWASFPQ-----VAGRR--- 298  
DB 488 AQQGVQHPVVRGADGPARRRRAPQPVDRD-----AAPRGAADDVQQRRAVPRRARG 543

QY 299 -----PDPPSPVPRGQGGQGWGTGTHG-PR-LEPDPQHFGAWLPLLSNGPHASSLW 352  
Db 544 PGAVPDVPRAPRGGR-EG--GLGARPRGPRVPRPEPR----- 581  
QY 353 SLFAPSPIPRCSGESEOLRACSQAPCPPEQDPDRALQCAAFNSQEPWGLYQWEPTEV 412  
Db 582 -----AAPPRAHPEA----- 592  
QY 413 QGSORCELNCRPRFRFVVRHTEKVDGTLTQCPGADICVAGRCLSPGCGDILGSGRRPD 472  
Db 593 -----PRARHRR-----AP-----RR-- 606  
QY 473 GCGVCGDDSTCLVSGNLT-----RGGLGYQKILWIPAGALRLQIAOLRPSN 523  
Db 607 -----VRLVGRLLRRARRALLRLRAGA-----FPAAGAVRLLRSPA 649  
QY 524 YIAL-RGPGRSII-----NGNAVDP-PGSYRAGGT--VFRYVRPREBEK----- 566  
Db 650 GVAVRGGPGRAPVLDDEALGRAAPGPRAGPRAGGAKDVARGAAEPRRRGRAHPDQGV 709  
QY 567 ---GESLSAEGPT---TOPVDVYMFQENPGVYQVVISPPP-----ILENTP 611  
Db 710 LRAGGDLPHGPARVQARPAH---RARGPRGHRR--REPEPRDRADRLLRQPAR 764  
QY 612 EPPVQLOPEILRVEPPLA-----PAPRPA-----RTPGTLQROVRIPQ 650  
Db 765 ADRLRGRPDVRRAGPRARRRARGVAPRGPAVRPARGPAAEGRRARAPLRREVRHAQ 824  
QY 651 MPAPP 655  
Db 825 RARPP 829

RESULT 6  
US-08-963-825-18  
; Sequence 18, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Bonde, Martin  
; APPLICANT: Qvist, Per  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1341 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (I)  
US-08-963-825-18  
  
Query Match 4.4%; Score 216.5; DB 3; Length 1341;  
Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
Matches 243; Conservative 62; Mismatches 411; Indels 345; Gaps 56;  
  
QY 22 CLQDEVLSGH-----SLOTPTTEG-----QGPGEVGMWQVNASCSQCGVGVQRRSRT 70  
Db 81 CPGAPEGECCVPCDGSSESTQDQETTVGEKPGKDTGPRGPRGAGPGGRDQIPQ--- 137  
QY 71 COLPTVOLHPSLPLPRPPRHP-----PGLPGPPGPPGPPGLGGNFAQLYDEKSTGGISVPGPMGPGSGRGL 188  
Db 138 -----PGLPGPPGPPGPPGPPGLGGNFAQLYDEKSTGGISVPGPMGPGSGRGL 188  
QY 100 QGPRPQTSPTL--PLYRTOSRGGPL--RGPAHLGRE-ETQETRAARRSLRDPKIP 154  
Db 189 PGPPGAPGPGZGFZPGZPGZPGZPGZPGZPGZPGZPGZPGZPGZPGZPGZPGZPGZPG 246  
QY 155 GMFCYGRVFPALPLHRRNRHPRSPRSELSSIRSEEAIPSPTPRAEFPFSAANGSPQTEL 214  
Db 247 GPZGARGLPGTAGLPGMKGH-----RGFSLBGAKG-BAGPA-GPKGZP---GSPGZBG 295  
QY 215 PPTLSVHTPSQ---AEPLSPET-----AQTEVAP---RTRPAPLHHHPRRAQASGE-- 261  
Db 296 APGZMG--PPGKNGSGEPGAGSGKGTGAKGPPGVGVQVQVQVQVQVQVQVQVQVQVQV 353  
QY 262 -----PPSPHSLGEGF-----FRASPPRRPSSQGWASQVAVR----- 297  
Db 354 TGLPCPPGERGCGSRFPFGADGAGVAGPKGAGRGSGPGAPGK---GSPGEAGRCEAG 409  
QY 298 -----RPPDFPVPRGRGQOQ-----QGWGTGTGTP 323  
Db 410 LPGAKLTGSPGSPGDKTGPAGQDGRPGP-PGPPGARGQAGVGMFPFGKGAAGEP 468  
QY 324 --HGPRLEPDQHPGAWLPLLSNGPHASSLWSLFPSPSPRCSGESEQLRACS----- 375  
Db 469 GKAGERVPGP--PCAVGPAGKDEAGAQ-----GPPGAPGAGERGEQGPAGSPGQGL 521  
QY 376 -----QAPCPPEQDPRAL-----QCAAFNSQEFMGQLYQWEPFTEVQGSQRCENCR 423  
Db 522 PGAPGPGCEAGKPGEGVPGDLGAPGPGSARGERFPGE-----RGVQGPFGAG 571  
QY 424 PRGFRFYVRHTEKYODGTLCQPGADICVAGRCLSPGCGDILG-----SGRRPDG 473  
Db 572 PRG-----ANGAPNDGAKGDAGAG--APGSQAPGLQMPGERGAGLPGKGRGD- 623  
QY 474 CGVCGDDSTCLVSGNLTDRGGPLGYQKILWIPAGAL--RLQIAQLRPSNLYALRPG 531  
Db 624 AGPKGADGSPGKDGVRGLTGPIGPPG-----PAGAPGDKGESGSPGAPGTGARGAPG 676  
QY 532 GRSIINGNAVDPGGSYRAGGTVFRYVRNPPREEG-----KGESLSA-----EGPT 576  
Db 677 DRG-----EPGPPGAGFAG-----PGADGQKAGPEGDAGAKGDAGPPCPGAPGA 723  
QY 577 TQPDVYMFQENPGVYQVVISPPPILNPTPE-----PPVQLOPEILRVEPPLA 632  
Db 724 GPPGPIGNV---GAPGAKGARGSPGATGFPGAAGRVGPPPS-----GNAGPPGPPG 775  
QY 633 PR-----PARTGTLQROVRLPOMAPPHPRTPLGSPAAYKWKVGHSAACS 679  
Db 776 PAGKEGKGPRGRTGTPAGRPGEV-----QPPGPPGAGKEKSPGA-----DGPAGAPGT 824  
QY 680 CG-----KGVWRPITFLCISRESGEELDERSCAGARPPASPEP 717



Db 825 PPGIAGQGRVGLPQGRGRRGPPGLPGP-----SGEPKQ--GPSGASGERGPPGPMG 877  
Qy 718 CHGTPCPPYWEAGWT-----SCSRSCGPGTQHRQLQCRQEFGGGSSVPPRCGHLPRP 772  
Db 878 PPLAGPESGREGAPGAEGSPGRDGSPP-----AKGDRGETGPPAGPPGAXGAP 930  
Qy 773 N-ITQSCQLRLCGHWEVGSVPSQSVRCGRQSRVRCVGNNGD--EVSEQ-----821  
Db 931 GPVGPAGKSGDRGETGPPAGPAGVPGVAGAGPAGPQ-----GPRDKGETGQDGRGKXG 986  
Qy 822 ---ECASGPPPPSREACDMGFCCTTAMFHSWSSKVSPEPP 859  
Db 987 RGFSGLQPPGPPSPG-EQGP-----SGASGAPGPRGP 1019  
RESULT 7  
US-09-570-573-18  
; Sequence 18, Application US/09570573  
; Patent No. 6342361  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/570,573  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1341 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (I)  
US-09-570-573-18  
Query Match 4.4%; Score 216.5; DB 4; Length 1341;  
Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
Matches 243; Conservative 62; Mismatches 411; Indels 345; Gaps 56;  
Qy 22 CLDQEVLSGH-----SLQTPTEG-----QGPESGVGPVWQVASCSPGCGVQVRRSRT 70

Db 81 CPGAEPVEGECPCVCDGSESPDQTTTVEGPKGDTGRGRGPRGAGPPGDRGIPQ-----137  
Qy 71 CQLPTVOLHPSLPLPRPRPH-----EALLPRG- 99  
Db 138 -----PGLPGPPGPPGPPGGLGNFAPQLVGYDEKSTGTGTSVPGMPSGPRGL 188  
Qy 100 QGRPQTSPEL--PLYRTQSRGRGGL--RGPASHLRE-ETQETRAARRRLRPIKP 154  
Db 189 PGPPGAPGPGZGFZGPGZPCZPCASGPMGPRGPPGPKGBGBGZAGKPRGZRGPP--P 246  
Qy 155 GMFGYGRVFPALPLHNRHRRSPRSELSLSSREEAIPSPTPRAEPFSANGSPQTEL 214  
Db 247 GPZGARGLEGTAGLPGMKH-----RGFSLBAGK-BAGPA-GPKGZP-----GSPGZG 295  
Qy 215 PPTLSVHTPSQ---AEPLSPET-----AQEVAP-----RTRPAPLRHHHPRAQAQSTE-- 261  
Db 296 APZMG---PPGKNGSEGPCAPGSKGDTGAKGEPGVGVQPPGPGAGEBGRGARGEPG 353  
Qy 262 -----PPSPHSLGEGGF-----FRASFPQPRRSPSSQGWASPOVAGR----- 297  
Db 354 TGLPGPPGERGGPSRGFPQADGVAGPKGPAGBGRSGPPGAPK-----GSPGEAGRGEGAG 409  
Qy 298 -----RDPFFSVPRGRQOQ-----QSPWGTGGTP 323  
Db 410 LPAKGLTSGSPGPDGKTGPPGAGQDGRPGP-PGPPGARGAGVMGPPGPKGAAGEP 468  
Qy 324 --HGPRLEPDQHPGAWLPLLSNGPHASSLSWLFAPSSPIPRCSGESEQLRACS----- 375  
Db 469 GKAGERGVGP--PGAVGPAGKDGAGAQ-----GPPGPAGPAGERGEGGPPGPGFQGL 521  
Qy 376 -----QAPPEQDPDRAL-----QCAAFNSOEFNGQLYQWEPTEVQSGSORCELNCR 423  
Db 522 PGAPGPEAGKPGEQCVPGDLGAPGSGARGERGFPGE-----RGVGGPPGAP 571  
Qy 424 PRGFRFYVRHTEKVQDGTLCQPGAPDICVAGRLSPSCDGLG-----SGRRPDG 473  
Db 572 PRG-----ANGAPNDGAKGDAGAP--APGSGAPGLOGMGERGAAGLPGPKGRGD- 623  
Qy 474 CGVGGDDSTCLVSGNLTDRGGPLGYQKILWIPAGAL--RLQIALRPPSSNYLALRGP 531  
Db 624 APKGADGSGKGVRLTGPIGPPG-----PAGAPGKSGSGSPGAPGTGARGAPG 676  
Qy 532 GRSIINGNWAVDPPGSYRAGGTVFYRNRPREG-----KGESLSA-----EGPT 576  
Db 677 DRG-----EPGPPGAPGAG-----PPGADQPGAKGEPGDAGAKGADAGPPGAPGA 723  
Qy 577 TQPDVYMFQENPNPVYQYVVISPPPILENTPTE-----PPVPLQLOPEILRVEPPLAPA 632  
Db 724 GPPGPIGNV---GAPGAKGARGSPGATGFGAAGRVGPPGPPS-----GNAGPPGPPG 775  
Qy 633 PR-----PARTPGLQROVRIPQMAPPPHPTPLGSPAAYWKRVHSHSACSAS 679  
Db 776 PAGKEGKGPRGETGPPAGPCEV-----GPPGPPGAPGKGSFGA-----DGPAGAPGT 824  
Qy 680 CG-----KGWVRPFLCISRESGEELDERSCAAGARPPASPEP 717  
Db 825 PPGIAGQGRVGLPQGRGRRGPPGLPGP-----SGEPKQ--GPSGASGERGPPGPMG 877  
Qy 718 CHGTPCPPYWEAGWT-----SCSRSCGPGTQHRQLQCRQEFGGGSSVPPRCGHLPRP 772  
Db 878 PPLAGPESGREGAPGAEGSPGRDGSPP-----AKGDRGETGPPAGPPGAXGAP 930  
Qy 773 N-ITQSCQLRLCGHWEVGSVPSQSVRCGRQSRVRCVGNNGD--EVSEQ-----821  
Db 931 GPVGPAGKSGDRGETGPPAGPAGVPGVAGAGPAGPQ-----GPRDKGETGQDGRGKXG 986  
Qy 822 ---ECASGPPPPSREACDMGFCCTTAMFHSWSSKVSPEPP 859  
Db 987 RGFSGLQPPGPPSPG-EQGP-----SGASGAPGPRGP 1019  
RESULT 8

US-09-548-608-18  
; Sequence 18, Application US/09548608  
; Patent No. 6355442

## GENERAL INFORMATION:

; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/548,608  
; FILING DATE:

## CLASSIFICATION:

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319

## FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (I)

US-09-548-608-18

Query Match 4.4%; Score 216.5; DB 4; Length 1341;  
Best Local Similarity 22.9%; Pred. No. 1.9e-06;  
Matches 243; Conservative 62; Mismatches 411; Indels 345; Gaps 56;

QY 22 CLDQEVLSGH-----SLQTPTEG---QGPEGVGWQVQWASCSQPCGVGVQRRSRT 70  
DB 81 CPAEVEGECPCVCDGSESFDQTTGVEGPKGDTGRGPRGAGPRGDIQO--- 137  
QY 71 COLPTVOLHPSLPLPRPRHP-----EALLPRG- 99  
DB 138 -----PGLPGPPGPPGPPGLGGNFAPQLYGYDEKSTGTGTSVPFGMPGSPRGL 188  
QY 100 QGPRQTSPTL--PLYRQSRGGL--RGPASHLGR--EQEIRARRSLRDPKIP 154  
DB 189 PGPPGAPPGGFTGGPCZPGZPGZPGASGPMRPPGPPGKBGGZAGKPGRGZRP--P 246  
QY 155 GMFGYGVFPFALPHNRHRRSPRSELSLSSRGEAIPSPTPRAEPFSANGSPQTEL 214  
DB 247 GPZGARGLPCTAGLPKMGH-----RGFSLBCAKG-BAGPA-GPKGZP-----GSPCZBG 295  
QY 215 PPTELSVHTSPQ---AEPLSPET-----AOTEVAP---RTRPAPLRHHPRACASGE-- 261

DB 296 APGZMG--PPGPKNSGEPGAPGSKGDTGAKGEPGVGVQPPPGPAGGCKRGARGEPGP 353  
QY 262 -----PPSPHSLGEGF-----FRASQPPRRPSSQGWASPOVAGR----- 297  
DB 354 TGLFGPPGGERGGPSRGFPFGADGVAGPKPAGERGSPGAPK-----GSPGAGRPGGAG 409  
QY 298 -----RPDPFSPVRCRGOOG-----OGPMGTGTGP 323  
DB 410 LPGAKGLTGCSPGPDGKTGPPGAGQDGRGP-PGPPGARGAGQVGMFPFGKAGAP 468  
QY 334 --HGPRLEPDPQHPGAMLLPSNGPHASSLSLFPAPSPIPRCSESEQLRACS----- 375  
DB 469 GKAGERGVGP--PGAVGPAGKDGGAQA-----GPPGAPGAGERGEGQGPAGSPGQGL 521  
QY 376 -----QAPCPPEQDPRAL-----QCAAFNSOFFMQLYQWEPFTEVQCSQRCNCR 423  
DB 522 PGAPGPGGAGKPGEGQGVPGDLGAPGSGARGERGFGE-----RGVQPPPGAG 571  
QY 424 PRGFRFYVRHTEKVDGTLCPQGAPDICVAGRCILSPGCDGILG-----SGRFPDG 473  
DB 572 PRG-----ANGAPNDGAKGDAGAP--APGSQAGPLQGMPPGERGAAGLPKPGDRGD- 623  
QY 474 GCVGGDDSTCRLYSGNLTDRGGLGYQKILWIIFAGAL--RLQIAQLRPSNYLALRPG 531  
DB 624 AGPKGADGSPGKGVRLTGTGIPPG-----PAGAPGDKGESGSPGPGAGTARGAPG 676  
QY 532 GRSTINGNVAVDPPGVSYRAGGTVERYNRPREEG-----KGESLSA-----EGPT 576  
DB 677 DRG-----EPGPPGAPGAG-----PPCADGPGKAGGPDAGAKGDAGPPGPAGPA 723  
QY 577 TOPVDVYMFQENPGVYQVVISPPPILENPTPE-----PPVQLQPEILRVEPPLAPA 632  
DB 724 GPPGPIGNV--GAPKAGARGAGPPGATGFPGAAGRVGPPGPS-----GNAGPPGPPG 775  
QY 633 PR-----PARTGTQORVRIQMPAPPHPHTPLGSPAPAAWKVHVSACSAS 679  
DB 776 PAGKEGKGPRGEGTGPAGPGEV-----GPPGPPGAGEKGSFGA---DGPAGAPT 824  
QY 680 CG-----KGWVRPIFLCISRESGSEELDERSCAAGARPPASPPEP 717  
DB 825 PGQGIAGQGVVGLPGQGRGPPGLPFP-----SGEPGKQ--GPSGASGERGPPGPMG 877  
QY 718 CHGTCPPPYWEAGWT-----SCSRSCGPGTQHRLOLQCRQEFGGGSSVPPERGHLPRP 772  
DB 878 PPGLAGPPGESGREGAPCAESGPDGSGF-----AKDRGETGTPAGPPGAXGAXGAP 930  
QY 773 N-ITQSCQLRLCGHWEVGSQSVRCRGQRGRQVRCVGNNGD--EVSEQ----- 821  
DB 931 GPVGPAGKSGDRGETGPAGPAGVPVGPAGARGPAGPQ-----GPRGDKGTGEGQDRGKNGH 986  
QY 822 ---ECASGPPPPSREACDMGPTATWFHSDWSKVSPEPP 859  
DB 987 RGFSGLQGPFGPPGSPG-EQGP-----SGASGAPGPRGP 1019

## RESULT 9

US-08-642-255-72  
; Sequence 72, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CAPPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 EHT UR
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1065 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-72

Query Match 4.4%; Score 213; DB 1; Length 1065;
Best Local Similarity 23.7%; Pred. No. 2.4e-06;
Matches 208; Conservative 32; Mismatches 341; Indels 296; Gaps 46;

QY 62 VGVQRS-----RTQQLPTVLHPSLDLPPRPHPEALLPRG-----QGRPOTSPTETLPL 113
Db 4 VVLRDWDENPGVTQNLRLAAHPTFASDPMGAHPAG--PKGAHPAGPKGAQGPAG--- 58

QY 114 YRTQSRGRGP--LRGPAHLGRETEQETRAARRSLRDPDKPKMGFGYGRVPFALPLHRN 171
Db 59 -PGSGRGDPGPAQGPAGPGG-----SR-GDPGPPGAQ----- 91

QY 172 RHRPRPPRSELSISSRGEERAIPTPRAEPFANGSPQTELPTELTSVHTPSQAEPL 231
Db 92 ---PAGFG-----GSRGDPGPPGAQGPAGPGSGRGDPG---PPGAQGPAGPGSGRDP 138

QY 232 SPETAETEVA---RTRAPLRRHRAQAGTE---PPSPTHSLGEGFFRASPQRRP 284
Db 139 GPPGAQGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQGPAGPGG--SRGDPG--P 195

QY 285 SSQGWASQVAGRRDPFVSFRG-----RQQG-----QGPWGTGTF----- 322
Db 196 GAQGPAGP--GSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQ 253

QY 323 ---PHGPRLEDPQHPGAWLPLLSNGPHA--SSLWSLFAFSPPIPRCSGESEQLRACSOAP 378
Db 254 PAGPGSGRGDPG--PGAHGAPGKAHPAGPKGAHPAGP--KGAHPAGPKGA--QGP 308

QY 379 CPP-----EQDPRALQCAAFNSQEFMGQLYQWEPFTEVQGSORCEINCRPRGFYVRHT 434
Db 309 AGPGSGRGDPGPPGAQGP-----GPGGSR----- 333

QY 435 EKVDGTLCPGAPDLCVAGRLS-----PGCDGILGSG-----RRPDGCGVCGG 479
Db 334 -----GDPGPPGAQGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQGPAGPGSG 388

QY 480 DSTCLVSGNLTDRGCGLYQKILWIPAGALRLQIAQLRPSNVIALRPGCGRSIINGN 539
Db 389 DP-----GPPGAQGP-----PAGP-----GSGRGDPGPPGAQGPAGPGSGRGD 425

QY 540 WAVDPGSTRAGTVRYNRPREKGSLSAEGTTPQVDVYIMTFQENPGVYQYVI 599
Db 426 --PGPPGAQGPAG-----PGSGRGDPGPPGAQGPAGPG-----GSRGDPG----- 463

QY 600 SSPPTILENPT-----PEPPVPO--LQPEILRVERPLAPRAPARTPTGLQQRVIP 649
Db 464 ----PPGAQGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQGPAG--CPGSGRGD----- 515

```

```

QY 550 QMPAPHPRTPLGSPAAYWKRVGHSACSACSGKGVWRPIFLCISRESGEELDBERSCAAGA 709
Db 516 --PGPPGAHGPAGKGAHPAGKGAHPAGKGAHPAGKGAHPAGKGAHPAGKGAHPAGKGAHP 557
QY 710 RPPASPEPCHGTFCPP---YWEAGEWTSCSRSCGPGTQHRQLQCRQEFGGGSSVPPPERCG 767
Db 558 QGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPPGAQ-----GPAGPGSGRGDP----- 606
QY 768 HLPNPITQSCQLRLOCHWEVGSWQSCSVRCGRGQSRQVRCVGNNGDEVSEBQECASGP 827
Db 607 -----GPPGAQGPAGPG-----GSRGD----- 626
QY 828 -----POPPSREACDMGPTTAMFHSWSSKVSPEPP 859
Db 627 PGAQGPAGPGSGRGDPGPPGAQGPAGPGSGRGDPGPP 663

RESULT 10
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
; US-08-931-820-1

```

```

Query Match 4.3%; Score 210.5; DB 3; Length 1057;
Best Local Similarity 23.6%; Pred. No. 3.5e-06;
Matches 235; Conservative 49; Mismatches 382; Indels 329; Gaps 55;

QY 36 PTEEGQ-----GPEGVWGPWQWASQPCGVGVQRRSNTCOLPTVQLHPSLPLPPRPRH 91
Db 146 PGERGRPCAPGAPARGNDGATGAAGPPGTG-----PAGEPGF 184

QY 92 PEALLPRG-----QGRPOTSPTETLPLYTQSRGRGP-----LRGPASHLGRRETQIRA 142
Db 185 PGAVKAGKAGPQGRSGEPQGV-----RGEPPGPPGAGAGPAGNPGADQPGKAG 237

QY 143 ARRSRLRDPKPGMFGYGRVPFALPLHNRHRRSP--PRSELSLI-----SSRGEAI 194
Db 238 ANGA-----PGTAGAPGFFGA-----RGSPGPGPGTGAAGLVGEPGAGSKSGS- 285

QY 195 PSPTPRAEPFANGSPQTELPTELTSVHTPSQAEPLSPETAQTEVAPRTAP--LRHH 252
Db 286 ---NKGEFGSA--GPQG--PP-----GPSGEGKRGPNCEAGSAGP---PGPPGLRG 328

```

```

QY 253 PRAQ-----ASGTEPPSPHSLGEGFFRASQPPRRSPGQWASPOVAGR----- 297
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 329 PGRGPPGADGVAGK-----GPAG-ERGSPPGPK-----GSPGEAGRPGEAGLEGARG 378
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 298 -----RPDPSPVPRGRGQOG-----QGPWGTGTGP--HGPR 327
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 379 LTGSPGSPGDKTTPPGAGQGRPGP--PGPPGARGQAGVMGFPKGAAGEPKAGER 437
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 328 LEPDPQHPGAWPLLSNGPHASLSLSPAPSPIPRCSGESQRLACS----- 375
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 438 GVPGP--PGAVGPAKGDEAGAQ-----GPPGPAAGRGEGQAGSPGFQGLPGPAGP 490
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 376 --QAPCPPOPDRAL-----QCAAFNSQFMQLQWBEFFVQGSORCELANCRGRGFR 429
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 491 PGEAGPGEQGVGDDLGAPGSPGARGERFPGE-----RGVGGPPGAGRG----- 537
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 430 YVRHTEKVDGTLCQAPADICVAGRCLSPGCDGILG-----SGRRPDGCGVCGG 479
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 538 --ANGAPNDGAKDAGAPG--APGSQAPGLQGMFGERGAAGLPGKGDGRD--AGPKGA 592
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 480 DDSTCRILVSNLTDGPGLOYKILWIPAGAL--RLQIAQLRPSSNYLALRPGGRSIIN 537
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 593 DGSPGDKGVRLTGPIGPPG-----PAGAPDKGESGSPGAGTARGAGPDGRG----- 642
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 538 GNAVADPPGSYRAGGTVFRYRPPREEG---KGESLSA-----EGPTIQVDV 582
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 643 ---EPGPPGAGTAG-----PPGADGQPGAKGEPGDAGAKGDAGPPGPAAGPPGPI 692
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 583 YMIQENPNQGVYQVYISPPPILENTPE----PPVQLQPEILRYEPPLAPAPR---- 634
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 693 GNV---CAPGAKGARGAGPPGATGPPGAGRVGPPGPPS-----GNAGPPGPPGAKGEG 744
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 635 -----PARTGTLQQRVRIQPPAPPHRTPLGSPAAVYKRWYGHGACSACSG---- 681
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 745 GKPRGTGTAGRGEV-----GPPGPPGAGEKSGFGA-----DGPAGATPQPGQI 793
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 682 -----KGVWRPIFICISRESGEELDERSCAAGARRPASPPEPCHGTTC 723
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 794 AGQGVVGLPQGRGERGFPGLPGP-----SGEPKQ--GPGSASGERGPPGMPGPPLAG 846
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 724 PPWEAEWNT-----SCRSRGCTQHRLQCRQEGGGSSVPPERCGLHLPEN-ITOS 777
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 847 PPGSGREGAPGAGSGPRGDSGP-----AKDGRGTGAPPPGAPGAPGAPGAPG 899
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 778 CQLRLCHWEYGSFWSQSVRCRGQRGRQSRQVRCVGNNGD--EYSEQ-----ECA 824
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 900 KSGDGRGTGAPGAPGAPGAPGAPGAPGQ-----GPRGDKGTGEQDGRGKGHRGFSGL 955
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 825 SGPPQPSRREACDMGCTTAWFHSMDNSKVSPEPP 859
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 956 QGPPGPPGSPG-EQGP-----SGASGAPGAPRGP 982
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

## RESULT 11

```

US-08-317-310A-64
; Sequence 64, Application US/08317310A
; Patent No. 3658701
; GENERAL INFORMATION:
; APPLICANT: WHITE, Morris F.
; APPLICANT: SUN, Xiao Jian
; APPLICANT: PIERCE, Jacalyn H.
; TITLE OF INVENTION: THE IRS FAMILY OF GENES
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,310A
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Louis Myers
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: JDP-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1321 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-317-310A-64

```

Query Match 4.3%; Score 210.5; DB 2; Length 1321;

Best Local Similarity 21.0%; Pred. No. 4.6e-06;

Matches 196; Conservative 79; Mismatches 313; Indels 345; Gaps 48;

```

QY 59 PCGVGVQRRSRTCQLPTVQLHP-----SLPLPPRRPHPEALLPRQG 101
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 417 PAG-GALQHSRSMSPVAHSPAAATSPGSLSSSGHSGSYPLPGSHPLPLHHPQG 475
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 102 PRPQT-----SPETLPLRYQSRGCGPL 125
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 476 QRPSSGASASGSDPGFMSLDYSGSPGDLRAFSSHRSNTPESIAETPPARDGGGEL 535
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 126 RG-----PASHLGR-----EETQETRAARRSLRDIKPMFGYGRVPFALPLHNR 172
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 536 YGMSMDRLSCGRPYRVSGDGAQLDLRGLKRTYSLTIPAR-----QRQV 583
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 173 RHPRPPRSELSLI-----SSRGEALPSTPRAEPFANGSQOTELPTELTVH--- 222
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 584 POPSSASLDEYTLMRATFSGSGRLCPSPASSPKV---AYNPYPE-DYGDIEIGSHKS 639
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 223 -----TSPQAEPLSPETAQTE---VAPRTRPAPLR-HHPRQA---SGT 260
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 640 SNNLGADDDYMWTFCAALRSQGNPSCKDDYMPMSPIVSAPKILQPLRLAALPPSGA 699
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 261 EPPSPHSLGE-----GGFFRASQPPRRPSQG---WASQVAGRRDPDPFSPVRGEG 310
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 700 AVPAPSPGVGRTFPVNGGYKASSPAESSPEDSGYMRMWCGLSMENPDP-KLLPNGDY 758
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 311 -QOQGGPWGTGTPHGPRLPDP-----POHPGAWPLLSNGPHASSLSWL 354
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 759 LNMSPSEAGTAGTP-----PDFSAALRGSGEGLKIPGHCYSSLP-----SY 801
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 355 FAPSSPIPRCSGESEQLRACQAP---CPPEQDPDPRALQCA-AFNSQEFMGOLYQWEPPT 410
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 802 KAPCS---CSGNDQYVLMSSPVGRILEERLEPQATPGAGTGA---AGGSTQPHHS 854
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 411 EVQGSQR-CELCNCRPRGFRFY---VRHTEKYQDG--TL-----CQGPAP----- 448
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 855 AVPSMRRPSAIGGRPEGFLGQRCAVRPRLSLEGLQTLPSMQEYPLPTEPKSPGEYIM 914
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 449 DICVAGRCILSPGCDGILGSGRRPDGCGVGDDSTCRLVSGNLTDGGLGKILWIFA 508
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 915 DFGEAGTRLSPAPPILLASAA-----SSSLLSAS-----SPA 947
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 509 GAL-----RLQIAQLRPSSNYLALRGCGRSIINGNNAVDPGPGYRAGGTVFRYN--- 558
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 948 SSLSGTPTGTTSSDRQSRSPSLDYMNLDFSPKS-----PKPSTRSGDTVGSMDGLL 998
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 559 -----RPPREKGKESLSAEGTTPQVDVYMI-----FQENPGV 593
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 999 SPEASSPYPPLPPRPPSTSPSSLIQQPLPPA-PGDLYRLPASAATSQGTAGSSMSSEPGD 1057
  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```
QY 594 FYQY-----VISSPP-PILENPTPE-----PPV 615
DB 1058 NGDYTEMAFGVAATPPQIVAPPKEGARVASPTGKRLKSLMDQVSGVEAFLOVSQPPD 1117
QY 616 PQLOPEILRVPEPLAPAPAPARTPCTLQVRIPQMPAPPHPRTPLGSPAPYMKRVGHSA 675
DB 1118 PHRGAKVIRADP---QGRRRHSHSETTSITTV-----TPV-SPSFAHNSKRHNS 1163
QY 676 CSASCKGWKWRPIFLICISREGEELDRSCAAGARPASP---EPCHGTFCPPPYWEAGEW 732
DB 1164 ASV-----ENVSLRKSSEGSTL-----GGGDEPPTSPGOAQL---VAVPPVFOARPW 1209
QY 733 TSCSRSCGGTQHROLOCRQEFGGGSSVPPER 765
DB 1210 -----NPGQCALIGC-----PGSSSPMR 1230

RESULT 12
US-07-609-716-65
; Sequence 65, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-609-716-65

Query Match 4.3%; Score 209; DB 1; Length 408;
Best Local Similarity 23.1%; Pred. No. 1.5e-06;
Matches 145; Conservative 14; Mismatches 200; Indels 270; Gaps 29;

QY 62 VGVORR-----RCQLETVOLHESL-----PLPPEPRHREALLPRGGGPRPQTS 107
DB 4 VVLQRRDWEVNGVYQLNRLAAHPFASDPMGAGPGPGPGPGPGAPGPGPG-PGPPGP 62
QY 108 PETPLRYTQSRGGGPGPLRGPASHLGRETOEIRAAARRSLRDPKPGMGYGRVFPALP 167
DB 63 P-----GPGFVGSFGA-----PGPPGPPG----- 82

QY 168 LHRNRHRSPPRSELSSIRGEEAIPSTPRAEPFSANGSPQTELPTELTSVHTPSQ 227
DB 83 -----PGPPP-----CAPGPPGPGPPGPPGAGPVGSPGAPGPP-----GPPGPP 123
QY 228 APLSPTAQTEVAPTRPAPLRRHHPRAQASGTEPSPTHSLGEGGFFRASQPPRPSQ 287
DB 124 GPPGAP-----GPPGPPGP-----PGPPGAGPVGSPG-----APGP---PGPP 159
QY 288 GWASPOVAGRRPDPFFSVPRGQQG-QGFWGTGTPHGRLEPDPQHFGANLPLLSNGP 346
DB 160 GPPGPGA---PGP-PCPPGPPGPGAGPVGSPGAPGPPGPGPGPCAGAPGPPGPPGP 215
QY 347 -----HASSLWSLFASSPPIPRCSGESEQLRACSAQCPEPQPDPRALQCAAFNSQEFMG 401
DB 216 PGPPGAGPVGSPGAPGPPG-----PGPPGPPGAPGPPG----- 250
QY 402 QLYQWEPFTEVQGSQRCNELNCRPRGRFRFYVRHTEKVDGTLCOQGAPDICVAGRCLSPGC 461
DB 251 -----PPG-----PPGPPG---PAGPVGSPGA 369
QY 462 DGLSGRRPDGCGVCGDDSTCLVSGNLTDGGLGVQKILWIPAGALRLQIAQLRPS 521
DB 270 PGPPGPPGPPGPPGAPGPP-----GPPG-----PPG----- 296
QY 522 SNYLALRGPGRSITINGNWAVDPPGSYRAGGTVRYNRPRRPEGKGESLSAEGTTQPVD 581
DB 297 -----PGPAGPVGSPGAPGPP-----PPGPPGAPGPPGPPGPP--- 333
QY 582 YMIFQENFGVYQYVVISPP-----PILENPTPEPPVQLOPEILRVPEPLAPAPAR 637
DB 334 -----GPPGAGPVGSPGAPGPPGPP-----GPPGAGPP----- 365
QY 638 TPGTLQVRIPQMPAPPHPRTPLGSPAA 666
DB 366 PGP-----PGPPGPPGAGPVGSPGA 386

RESULT 13
US-08-475-411A-65
; Sequence 65, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A 55186--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 408 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PS-08-475-411A-65

Query Match 4.3%; Score 209; DB 4; Length 408;  
Best Local Similarity 23.1%; Pred. No. 1.5e-06;  
Matches 145; Conservative 14; Mismatches 200; Indels 270; Gaps 29;

Qy	62	VGQVRRS----	RTCQLPTVOLHPSL-----	PLPRPRRHPEALLP	QCGGRPQTS	107	
Db	4	VLQRDOWENPGVTQNLNRLAAHSPFASDPMGAPGPPGPGPGPGAGPGPG--	PGPGPG	62			
Qy	108	PETLPLYRTQSRGRGGPLRGPASHLGRREETQETIRAAARRSLRDP	IKPGMFGYGRVPFALP	167			
Db	63	P-----	GPAGVPGSPGA-----	PGPGPGG-----	82		
Qy	168	LHRNRHPRSPRSELSLSRGEAEI	PSPTPRAEBSANGSPOTELPTELSVHTPSPQ	227			
Db	83	-----	PPGGP-----	GAPGPPGPPGPPGAGVPGSPGACGPP-----	GPPGPP	123	
Qy	228	AEPLSPETAOTEAVPTRAPLARHHPRAQA	SGTEPPSTHLSGEGGFFRASPQRRPSSQ	287			
Db	124	GPPGAP-----	GPPGPPG-----	PGPGPAGVPGSPG-----	AFGP--PGPP	159	
Qy	288	GWASPVAGRRPDPFVS	PRGRQOQ--OGPWGTGGTHPGHLEPDPQHPGAWLPLLSNCP	346			
Db	160	GPPGPPGA---	PGP--PDPGPPGPPGAGVPGSGRAGPPGPGPGPGAGCPGPGPPG	215			
Qy	347	-----	HASSLSLFPASSPI	PRCGESEQLRACSAQPCPPEQDDPRALOCANFNSOEFMG	401		
Db	216	PGPGPAGVPGSPGAPGPPG-----	PGPGPPGAGPPG-----	PGPGPPGAGPPG-----	250		
Qy	402	QLYQWEPTEYQGSORCELNC	RRGRFRFYVHRHTEKVODGTL	CGQFAGP	ADICVAGRCLSPGC	461	
Db	251	-----	PG-----	PGPG-----	PAGVPGSPGA	369	
Qy	462	DGILSGRRPDCG	VGCGDDSTCRLVSGNITDRGGPLGYOKI	LIWIPAGALRLQIALRPS	521		
Db	270	PGPPGPPGPPGAGPP-----	PGPP-----	PGPP-----	296		
Qy	522	SNYLALRPGGRS	IINGWAVDPPGSYRAGT	VFYRNRP	PRRKEGKESLSAEGTTQ	581	
Db	297	-----	PGPAGVPGSPGAPGPP-----	PGPGPPGAGPPGPPGPP-----	333		
Qy	582	VYMTQENPGVYQYV	ISSP-----	P	ILENPT	EPBPVQOLQELRVEGPLAPARPAR	637
Db	334	-----	GGPGPAGVPGSPGAPGPPGPP-----	GGPGAGP-----	GPPGAGP-----	365	
Qy	638	TPGTLQORQVRIPQ	MAPPPHRTPLGSPAA	666			
Db	366	PGP-----	PGPGPPGAGVPGSPGA	386			

RESULT 14  
US-08-478-029A-65  
; sequence 65, Application US/08478029A  
; Patent No. 6184348  
; GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A.  
 APPLICANT: Cappello, Joseph  
 TITLE OF INVENTION: Functional Recombinantly Prepared  
 TITLE OF INVENTION: Synthetic Protein Polymer  
 NUMBER OF SEQUENCES: 119  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 STREET: Four Embarcadero Center, Suite 3400  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: US  
 ZIP: 94111  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/478,029A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/609,716  
 FILING DATE: 06-NOV-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/269,429  
 FILING DATE: 09-NOV-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/114,618  
 FILING DATE: 29-OCT-1987  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 06/927,258  
 FILING DATE: 04-NOV-1986  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Tegcartin, Richard F.  
 REGISTRATION NUMBER: 31,801  
 REFERENCE/DOCKET NUMBER: A-55186-8/RFT/NTK  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-781-1989  
 TELEFAX: 415-398-3249  
 INFORMATION FOR SEQ ID NO: 65:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 408 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 08-478-029A-65

Query Match	4.3%	Score 209;	DB 4;	Length 408;
Best Local Similarity	23.1%	Pred. NO. 1.5e-06;		
14. Mismatches	200.	Indels	270.	Gaps

QY	62	VGQRRS	-----RTCQLTTVOLHPSL	-----PLPPRRPHREALLPQQGRRPQTS	107	
		:	:	:		
		:	:	:		
		:	:	:		
Db	4	VVLQRDWNENPGVTQNLRLAAHSPADSPMGAPGPGPGPGPGAGPGGP-PGPPGP	62			
		:	:	:		
QY	108	PETILPLYRTQSRGGGPLRGPSHLCREETQETIRAAARRSLRDP1KPGMFGYGRVPFALP	167			
		:	:	:		
		:	:	:		
Db	63	P-----GPAGVVGSPGA	-----PGPPGPG	-----	82	
		:	:	:		
QY	168	LHRNRHRSPPRSEULISRGREAI	PSPTPRAEPFSANGSPQNELPTELTVHPTSPQ	227		
		:	:	:		
		:	:	:		
Db	83	-----PPGPP	-----GAPGPPGPGPGPGAGVVGSPGAGP	-----GPGGP	123	
		:	:	:		
QY	228	AEPLSPETAQTEAVPRTRPAPLRHHRAQAAGTGPSPTHSLGEGGFFRASQPGRPSQ	287			
		:	:	:		
Db	124	GPGAP	-----GPPGPPG	-----PGPGGAGVVGSPG	-----APGP--PGPP	159
		:	:	:		
QY	288	GWASQVAGRRAPDPFVPRGRQOQ	-QGPGWGGTTPHGRLEPDPQHPGCAWLPILLSNGP	346		
		:	:	:		
Db	160	GPQGP	PPGA-----PGP-PGP	PGPGPGPGPGAGVVGSPGAPGPGPGPGPGPGPGPGPGPGP	215	

; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21

Query Match 4.2%; Score 205.5; DB 3; Length 1078;  
Best Local Similarity 22.5%; Pred. No. 7.8e-06;  
Matches 206; Conservative 48; Mismatches 286; Indels 375; Gaps 52;

[illegible]

AS 10 01 17 10 17 10

WED JUL 24 11:39:20 2002

Search completed: July 23, 2002, 21:05:25  
Job time: 2854 sec





OM of: US-10-041-770-2 to: Issued\_Patents\_NA.\* out\_format : pfs  
Date: Jul 23, 2002 9:42 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL-framet\_p2n.model -Dev=xlh  
-Q/cgn2\_1/USPIO.spool/US10041770/runat\_22072002\_153136\_2701/app\_query.fasta\_1.947  
-DB-Issued\_Patents\_NA -CFMT=fastp -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS-human40 cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10041770.ecgn1\_1\_59 -NCPG=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-10-041-770-2

Query length: 877

Database: Issued\_Patents\_NA.\*

Database sequences: 383533

Database length: 122816752

Search time (sec): 54.460000

## score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7		368.50	231.72	3,2e-05	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1		368.50	231.72	3,2e-05	44377
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-945-283-1		368.00	243.28	7,3e-06	8438
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1		359.50	192.73	0,0048	4411529
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-1		355.00	230.01	4,0e-05	15872
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-306-296B-3		347.00	234.52	2,2e-05	3991
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-320-878-19		347.50	217.28	0,0002	38506
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2		344.50	182.65	0,0173	4403765
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-13		342.00	222.18	0,0001	13987
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-7		342.00	213.91	0,0003	44377
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1		342.00	213.91	0,0003	44377
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-32		341.50	223.42	9,3e-05	11220
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-5		341.50	214.92	0,0003	36778
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11		341.00	222.60	0,0001	12001
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11		338.00	220.59	0,0001	12001
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-428-517-1		337.50	209.89	0,0005	50937
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-690-473-1		336.50	227.00	5,9e-05	4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-821A-1		336.50	227.00	5,9e-05	4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-659-1		336.50	227.00	5,9e-05	4257
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-384-490-30		335.50	229.75	4,1e-05	2640
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-459-383-30		335.50	229.75	4,1e-05	2640
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-537-002A-5		334.50	226.86	6,0e-05	3600
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-863-010-5		334.50	226.86	6,0e-05	3600
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-024-429-5		334.50	226.86	6,0e-05	3600
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-3		331.00	214.98	0,0003	13613
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-125-468-1		329.00	207.97	0,0007	30001
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-474-933-1		329.00	207.97	0,0007	30001
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-773-816-1		329.50	208.66	0,0006	23673
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-335-409-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-102-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-567-969-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-480-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-486-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-568-472-1		327.50	201.02	0,0016	68750
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-105-537-36		325.00	219.64	0,0002	4041
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-804-227C-1		323.50	201.65	0,0015	43280
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1		323.50	219.98	0,0001	3048
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-14418-1		322.50	219.98	0,0001	3048
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-15327-1		322.50	219.98	0,0001	3048
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-105-537-30		322.00	208.81	0,0006	13842
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-977-434-9		320.50	220.04	0,0001	2505

/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-458-819-9 + 320.50 220.04 0.0001 2505  
/cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:PCT-US91-07035-9 + 320.50 220.04 0.0001 25  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-643-734C-3 - 319.50 204.41 0.0011 20235  
/cgn2\_6/ptodata/2/ina/6A\_COMB.seq:US-08-439-009A-3 - 319.50 204.41 0.0011 20235  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-1 - 319.50 198.96 0.0021 43280

seq\_name: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-7

## seq documentation block:

Sequence 7, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

alignment\_scores:  
Quality: 368.50 Length: 1067  
Ratio: 0.899 Gaps: 59  
Percent Similarity: 38.425 Percent Identity: 23.524

alignment\_block:  
US-10-041-770-2 x US-08-804-227C-7

Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377

```

53 TrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgSerAr 69
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29210 TGGCGCGGTGCTGCTGGCGGAC...ACGGGTGCTCGCCCGTCCG 29256

69 gThrCysGlnLeuProThrValGlnLeuHisProSerLeuProLeuProp 86
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29257 GACGGTGGT...CCFAG...CCAGCCCGC 29282

86 roArgProProArg.HisProGluAlaLeuLeuProArgGlyGlnGlyPr 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29283 CTCGGTCCCGCGGTGTCTCTACGCGCTACTACGGTCTGGTCACTGG 29332

102 oArgProGlnThrSerProGlnThrLeuProLeuThrArgThrGlnSerA 119
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29333 CCGGCTCAGCGCGGGAGCGGGTCTGTCACCGCCCGCGGGGGC 29382

119 rg.....GlyArgGlyGly.....ProLeuArgGlyProAla 129
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29383 GTGGCATGCGCGGTGTCAGATCGCGCCACCTCGCGCGGAGTGTCT 29432

130 SerHisLeuGlyArgGluGluThrGlnGluLeuArgAlaAlaArgArgse 146
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29433 GGGCAC.....CGCGAGCCCGCAAGTGGGACGCCCTCGCGCCATGG 29476

146 rArgLeuArgAspProIleLysProGlyMetPheGlyTyArgValP 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29477 GCATCAGCGAGCA...C...C 29493

163 roPheAlaLeuPro.....LeuHisArgAsnArg 172
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29494 CTCGCCCTCTCCCGCACCTCGACTTCGCGACCGCCTTACCGGAGCGGA 29543

173 ArgHisProArgSerProProArgSerGluLeuSerLeuIleSerSerAr 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29544 CGGCACGTCCCGCGCGAGCTGCTCTGTAAGTCTGCTCACCAGGAGTTCG 29593

189 ggGlyGluAlaIleProSerProThrProArgAlaGluProPheSerA 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29594 TGGAGC.....CTCCCTCGGGGTGCTCGTCCGGCGCGGCTTCCTG 29637

206 laAsnGlySerProGlnThrCluLeuProProThrGluLeuSerValHis 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29638 GAGTGGCAGACGACGTCGCGGAGCCCGAGCGGATCCCGC...CGA 29684

223 ThrProSerProGlnAlaGluProLeuSerProGluThrAlaGlnThrG 239
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29685 ACACCGGGGTGCGTACCGGGGCTTCGACCTCAACGAGCGCGGACCG 29734

239 uValAlaProArgThrArgProAla..... 247
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29735 A.....CGCACTCGCGCGGTGTCGCGGAACTGATGGACCTGTC 29775

248 .....ProLeuArgHisHisProArgAlaGlnAla 257
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29776 GCGCGCGGTGCTGACCGCTGCGGTGCTGTCACCCACGCTGCGCG 29825

258 SerGlyThrGluPro.....ProSerProThrHisSerLeu... 269
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29826 GCGCGGACGACCCCTGCGCACTACGACCGCGCGGACCGGAAAGC 29875

270 .....GlyGluGlyG 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29876 TGTCTGTACCATGCGCGCGCTGGCACCGGTACGGCACGGTCTCTGTC 29925

273 lypPheArgAlaSerPro...GlnProArgArgProSerSerGlnGly 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29926 ACCGGTGGCAGCGCGCTGCGGAGCGGATCGCCCG..... 29963

289 TrpAlaSerProGlnValAlaGlyArgArgPro.....AspProPh 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
29964 .....CCACCTGGCGAGCGCGGCTCTCGCGGTGCTGATCGCG 30007

```

```

302 eProSerValProArgGlyArgGlyGlnGlnGly...Pro..... 316
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30008 CCGCGCGGGCCCGGACGCGGAGGCGCGCGGAGCTGGTCCGCGACCTC 30057

316 ..... 316

30058 GCGGCGCTGGCGGTGCGCCACCGTGTCTGCTGCGCTGCGAGTCTCCGACG 30107

316 ..... 316

30108 GGACCGCTCGCGGAGCTGCTCGCGCGCATACCGGCGCATCACCGGTGA 30157

317 .....TrpGlyThrGly.....G 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30158 CGCGGTGTGTCACAGCACCGCGGCTCTCGACGACGCGGTGCTGCGCGG 30207

321 yThrProHis.....GlyProArgLeuGluProA 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30208 CTCACCCCGAGCGGATCGCGCGGTGCTGCGGCCAACAGTGGAGGCGG. 30256

331 spProGlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHis 347
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30257 .....CCGTCCACCTGG.....ACGAACCTCACCC... 30280

348 AlaSerSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCy 364
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30281 GCGACCTCGACCTGTCGCGGTGCTCTCTCTCTCCAGCGCGGTCTG 30330

364 sSerGlyGluSerGluLeuArgAlaCysSerGlnAlaProCysProp 381
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30331 C.....TGGGAGCGCGCGCCGACGACGCGCGCGGCGGCGGCGGCG 30374

381 roGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGln 397
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30375 CCTCGACGCGCTCGCG..... 30391

398 GluPheMetGlyGlnLeuTyrglnTrpGluProPheThrGluValGlnG 414
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30392 .....CCGCGCGCGGTGCTCTCGCG 30411

414 ySerGlnArgCysGluLeuAsn.....CysArgProArgGlyPheArgP 429
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30412 CTCGCGGTGCTCACTCGCTCGGTGCTGCTGCTCGACACCGCGCGGAT 30461

429 heTyrgValArgHisThrGluLysValGlnAspGlyThrLeu..... 442
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30462 GGCACACGACGTGACAGGAGGCTCCAGCGCGCTTCGCGCGCGACG 30511

443 .....CysGlnPro...GlyAlaProAspIleCysValAlaG 454
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30512 GCTTCCCGCGCTGTCGCGCACGCTGGGCGCG..... 30544

454 yArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgP 471
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30545 .CGCTTTCGACGCGCGCTGC...GGGTGACGAGCGCGGTGCGAGTCCC 30590

471 roAspGly.....CysGlyValCysGlyGlyAspAspser 482
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30591 CATCGGTTCGACCGCGCGCTGCGCGCCACCGGAGCGTCCCGCGCC 30640

483 ThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuG 499
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30641 TGCTGTGCGACCTGCTCGGTCCG...CCCGCGCGACCGGTCCG..... 30682

499 yTyrgLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleA 516
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30683 .....C 30683

516 laGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGly 532
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
30684 GGCGCGCGGTGCGCGCGCTTCCGCTCGGACGCGCGGACCGGTGCGG 30733

```

```

533 ArgSerllelelsanglysnTrpAlaValAspProProGlySerTyAr 549
30734 AGCCGCTCGCGGAGCGGT...TGGCGGGACTCTCGCGCGAGAACGCCAC 30780
549 gAlaclyGlyThrValPheArgTyrAsnArg.....559
30781 GACCGGTCTCGCGCTGGTTCGCGGACACAGTGGCGCGGTACTGGGCCA 30830
560 .....ProArgGluGluGlyGlySerLeuSerAlaGlu 573
30831 CGGCTCGCGCGGAGTCCGGCGCGACCGCGCGGTTCGCGAGGTCGGGT 30880
574 GlyProThrThrGlnProValAspValTyrMetIlePheGlnGluAs 590
30881 TCGACTCGCTACGCGCGGTGACTCGGCAACCGGATGCGG...CGGTC 30927
590 nProGlyValPheTyrGlnTyrValIleSerSerProProIleLeuG 607
30928 ACCGGG.....TCAGGCTCCCGCCACCGCTGTCT 30959
607 luAsnProThrProGluProValProGlnLeuGlnProGluIle... 622
30960 CGACACCCACCGCGCGGTCTCTCGACCTCGACGCGCTGCTGG 31009
622 ..... 622
31010 CCCCGCACAGCGGTACACACACACCGGTCTGTCGCAACTGGACCGC 31059
623 ...LeuArgValGluProProLeuAlaProAlaProArgProAla... 636
31060 ATCAGGAGGCGCTCGCGCGCTCACCGCGACCGCTCGCGAGCTCGC 31109
637 .....ArgThrProGlyThrLeuGlnArgGlnValArgIleProG 650
31110 CCCCGCCCGACGACCGGCGCA.....GG 31135
650 nMetProAlaProHisProArgThrProLeuGlySerProAla... 665
31136 TCGCCCTCGCTGGAGCCCTGCGCGCGCTGCGCGCGCTGCGACGAC 31185
665 ..... 665
31186 GCGCGCGCGCGCGAGCGAGCATCACCGAGTCTGAGCAGCGCCGA 31235
666 .....AlaTyrTrpLysArgValGlyHisSerAlaC 676
31236 CGACGACGAGATCTTCGGTTCATCGAGCGAGCGGTACGCGACGCTG 31285
676 sSer.....AlaSerCysGlyLysGlyVal.TrpArg 686
31286 CGCGCGCGCGAGCGCGCGCTCATCAAGGAAGAACCCACCATGGCGA 31335
686 ..... 686
31336 AGAAGAGAGCTGGCGCGCTACCTCAAGCGCGTGCAGCGGTGAGTGCAC 31385
687 ...ProIlePheLeuCysIleSerArgGluSerGlyGluLeuAspG 702
31386 CGGGCCACCGAGCAGTCTCGTGGCTTGAGCGCGGCGCGCCACGAGCCGAT 31435
702 luArgSerCysAlaAlaGlyAlaArgProAlaSerProGlu..... 716
31436 CGCGATCTCGGGCGCGCTCCCGAGCTCCCGCGCGGTGAGAGTCCGG 31485
717 ...ProCysHisGlyThrProCys..... 723
31486 AGCACTGT...GGGAGTGTGTCACCGCGGTGCGGAGCGGTGCGCCCG 31532
724 .ProProTyrTrpGluAlaGlyGluTyrThrSerCysSerArgSerCys. 739
31533 GCGCGCGCGCGCGCGGTGGAGGTGAGGAGAGTACTCGCGCGACCC 31582
740 .....GlvProGlyThrGlnHisArgGlnLeuGlnCysArgGln 752
31583 CGACAGCGCGCGCACCTCGTACTGCGCGGAGCGCGGTTCGTGCGAGG... 31622
753 GluPheGlyGlyGlySerValProProGluArgCysGlyHisLe 769
31630 .....GGCGCGACCGGTTCGACCCCGCGCTTCTGCGCATCTCGCC 31670
769 uproArgProAsn.....IleThrGlnSerCysGlnLeuArgLeuC 783
31671 AACGAGCGCTCACCATGAGACCCCGAGCGGTGCTGCTGGAGACCTC 31720
783 ysgly...HisTrpGlu.....ValGlySerProTrpSerGln 794
31721 CTGGGAGCGGTGGAGCGCGGTGACCCCGCAGTCCCTGCGCGGCA 31770
795 CysSerValArgCys.....GlyArgGlyGlnArgSerArgGlnVal 809
31771 GCGGACCGCGCTGTTCGCGCGCGGTGGAGCGGTACCGAAGGCGC 31820
809 gCysValGlyAsn.....AsnGlyAspGluValSerG 820
31821 GTCGAGGCGCTCGAAGCGCATCTGGAGGCCCACTCTGCGCGCATCGT 31870
820 luGlnGluCysAlaSerGlyProGlnPro.....ProSerArg 833
31871 CAGCTTCACCGCGCGCGGTGCGCTAGCGCTGGCGCTGGAGGCGCGG 31920
834 GluAlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTr 850
31921 CGCTGAGATCGACAGCGCTCTCTCTGCTGCTGGGCGACTGCACCTG 31970
850 p..... 850
31971 GCGGTGCTGCTGCGCGCGGCGAGTGCACCTGCGACTGCGCGGCGG 32020
851 .....SerSerLysValSerProGluProPro 859
32021 CGCCAGGTGTCGCGGACTTCGCGCTTTCACCGAGTCTCC 32064
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-198-1
seq_documentation_block:
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:

```

```
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

alignment_scores:
Quality: 368.50 Length: 1067
Ratio: 0.899 Gaps: 59
Percent Similarity: 38.425 Percent Identity: 23.524

alignment_block:
US-10-041-770-2 x US-08-804-198-1 ..
Align seg 1/1 to: US-08-804-198-1 from: 1 to: 44377

53 TrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgSerAr 69
|||||
29210 TGGGCCCGGTGTCGGCGGACC...ACCGGTGTCGCCCGCCCG 29256
|||||

69 gThrCysGlnLeuProThrValGlnLeuHisProSerLeuProLeuProp 86
|||||
29257 GACGGTGT...CCTACG.....CCAGGCGCG 29282
|||||

86 roArgProProArg.HisProGluAlaLeuLeuProArgGlyGlnGlyPr 102
|||||
29283 CTGGTCCCGCGGTTCCTCAGCCCTACTACGGTCTGGTCAACCTGG 29332
|||||

102 oArgProGlnThrSerProGluThrLeuProLeuTyrArgThrGlnSerA 119
|||||
29333 CCGGCTCAGCGCGGAGAGCGGTGCTGTGCACGCCCGCGCGGCG 29382
|||||

119 rg.....GlyArgGlyGly.....ProLeuArgGlyProAla 129
|||||
29383 GTCGGATGCCCGCGGTGATGATCGCCCGCCCTCGCGCGGAGGTGT 29432
|||||

130 SerHisLeuGlyArgGluGluThrGlnGluIleArgAlaAlaArgSe 146
|||||
29433 GGCCAC.....CGCAGCCCGCAAGTGGGAGCGCCCTCGCGCCATGG 29476
|||||

146 rArgLeuArgAspProIleLysProGlyMetPheGlyTyrGlyArgValp 163
|||||
29477 GCATCACCGACGACCA.....C 29493
|||||

163 roPheAlaLeuPro.....LeuHisArgAsnArg 172
|||||
29494 CTGGCTCTCTCCCGCACCCCTCGACTTCGCGACGCCCTTACCGGAGCGGA 29543
|||||

173 ArgHisProArgSerProProArgSerGluLeuSerLeuSerSerAr 189
|||||
29544 CGCACGTCGCCGCGGACCTGCTGCTGAACCTGCCTACCAAGGATTCG 29593
|||||

189 gGlyGluGluAlaIleProSerProThrProArgAlaGluProPheSerA 206
|||||
```

```
|||||
29594 TGGACGC.....CTCCCTCGGCTGCTCGTCCGGCGGCGGCTCCCTG 29637
|||||
206 laAsnGlySerProGlnThrGluLeuProProThrGluLeuSerValHis 222
|||||
29638 GAGCTGGCAAGACCGAGCTCCGGGACCCCGAGGATCGCCGC...CGA 29684
|||||
223 ThrProSerProGlnAlaGluProLeuSerProGluThrAlaGlnThrGl 239
|||||
29685 ACACCCGGGGTGGCTACCGGGGCTTCGACCTCAACGAGCGGACCG 29734
|||||
239 uValAlaProArgThrArgProAla..... 247
|||||
29735 A.....CGCACTCGGCGGCTGCTGGGGAACGTGACCTGTTTC 29775
|||||
248 .....ProLeuArgHisHisProArgAlaGlnAla 257
|||||
29776 GCGCGCGGCTGTCACCCGCTGCGCTGTCACCCACGACGTGGCCG 29825
|||||
258 SerGlyThrGluPro.....ProSerProThrHisSerLeu... 269
|||||
29826 GGCGCGGACGCCCTGCGCACCATACCCAGGCCGCGCACCGGAAAGC 29875
|||||
270 .....GlyGluGlyG 273
|||||
29876 TCGTCTGACCATGCCCGCGCTGGCACCCGTACGCGACGCTCTGTC 29925
|||||
273 lyPhePheArgAlaSerPro...GlnProArgArgProSerSerGlnGly 288
|||||
29926 ACCGGTGGCACCGCGCCCTCGGACGCGCATCGCCG..... 29963
|||||
289 TrpAlaSerProGlnValAlaGlyArgArgPro.....AspProPh 302
|||||
29964 .....CCACCTGGCGAGCGCGCACGCGCTCGCGCGCTGCTGATCGCG 30007
|||||
302 eProSerValProArgGlyArgGlyGlnGlnGlyGlnGly..Pro..... 316
|||||
30008 CCGCGCGGCGCGCGGAGCGGAGCGCGCGAGCTGGTGGCGCGACTC 30057
|||||
316 ..... 316
|||||
30058 GCGGCCCTGGCGCGTGGCCACCGTGGTGGCTGGCGACTCTCCGACGC 30107
|||||
316 ..... 316
|||||
30108 GGAGCGCTCGCGGACTGCTCGCCGCATACCGCGCGCATCACCGCTGA 30157
|||||
317 .....TrpGlyThrGly.....Gly 321
|||||
30158 CGCGGTGTGCACAGACCGCGCTCTCGACGACGCGCTGCTCCCGGG 30207
|||||
321 yThrProHis.....GlyProArgLeuGluProA 331
|||||
30208 CTCACCCCGAGCGGATCGCGCGGTGCTGGGCCCAAGGTGGAGCGCG. 30256
|||||
331 spProGlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHis 347
|||||
30257 .....CCGTCACCTGG.....ACGAACTCACCC... 30280
|||||
348 AlaSerSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCy 364
|||||
30281 GCGACCTCGACCTGTCGGCGTTCGCTCTCTCTCTCTCCAGCGCGGCTG 30330
|||||
364 sSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProp 381
|||||
30331 C.....TGGGACGCCCGCGGAGCGCAACTACCGCGCGGCGCAC 30374
|||||
381 roGluGlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGln 397
|||||
30375 CCTCAGCGCCCTCGCG..... 30391
|||||
398 GluPheMetGlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGl 414
|||||
```

30392 .....CCGGGCGCGGTCCCTCGGC 30411  
414 ySerGlnArgCysGluLeuAsn.....CysArgProArgGlyPheArgp 429  
:|||||:|||||:|  
30412 CTCCTCGGTGTCACTCGCTGGGTCTGTGTCTCGACACACCGCGAT 30461  
429 heTyrValArgHisThrGluLysValGlnAspGlyThrLeu.....442  
:|||||:|||||:|||||:|||||:|  
30462 GGCACAGCACTGGACAGAGAGCCTCCAGCGGCTTCGCCCGCAGCG 30511  
443 .....CysGlnPro...GlyAlaProAspIleCysValAlaG1 454  
:|||||:|||||:|||||:|  
30512 GCTTCCCGCCCTGTCCGCCACGCTGGGGCGG.....30544  
454 yArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgp 471  
:|||||:|||||:|  
30545 .CGCTGTTCAGCCGCCCTGC...GGGTGACGAGCGCGTGCAGGTCCC 30590  
471 roAspGly.....CysGlyValCysGlyGlyAspAspSer 482  
:|||||:|||||:|||||:|  
30591 CATCGGTTCCGACCGCGCGCTGCGCGCCACCGGAAGCTCCCGGCC 30640  
483 ThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuG1 499  
:|||||:|||||:|  
30641 TGCTGTCGACCTCGTCTGGGTCCG...CCCGGCGACCGGTCCG.....30682  
499 yTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleA 516  
30683 .....C 30683  
516 laGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyGly 532  
:|||||:|||||:|  
30684 GGCCCGCGCTCGGCCCTTCCTCGGCTCGGACGCGCGGACCGTGGCG 30733  
533 ArgSerIleIleAsnGlyAsnTrpAlaValAspProProGlySerTyrAr 549  
:|||||:|||||:|  
30734 AGCGCTCGCGGAGCGGT...TGGCGGAGCTCTCCCGGAGAACGCCAC 30780  
549 gAlaGlyGlyThrValPheArgTyrAsnArg.....559  
:|||||:|||||:|  
30781 GACGGCTGCTCGGCTGTGTGCGGACACAGTGGCGCGCTACTGGGCCA 30830  
560 .....ProProArgGluGluGlyGlyGlyGluSerLeuSerAlaGlu 573  
:|||||:|||||:|  
30831 CGGCTCGCGCGGAGTCCGCGCGACCGCGGTTCGCGAGTGGGT 30880  
574 GlyProThrThrGlnProValAspValTyrMetIlePheGlnGluAs 590  
:|||||:|||||:|  
30881 TCAGCTCGTACGCGCGGTGGAACTGGGCACCGGATGGCG...CGGTC 30927  
590 nProGlyValPheTyrGlnTyrValIleSerSerProProIleLeuG 607  
:|||||:|||||:|  
30928 ACCGGG.....TCAGGCTCCCGCGCACCTGGTCTT 30959  
607 luAsnProThrProGluProProValProGlnLeuGlnProGluIle... 622  
:|||||:|||||:|  
30960 CGACCCACCCCGCGCGCTGTCTCTGCGACCTCGAGCGGCTGTGG 31009  
622 ..... 622  
31010 CCGCGCACAGCGGTACACACACACCGGTGTCTCGGAACCTGGACCG 31059  
623 ....LeuArgValGluProProLeuAlaPro..AlaProArgProAla... 636  
:|||||:|||||:|  
31060 ATCAGAGAGGCGCTGGCGGCTTCACCCGAGCACCTCGCGAGCTCGC 31109  
637 .....ArgThrProGlyThrLeuGlnArgGlnValArgIleProG1 650  
:|||||:|||||:|  
31110 CCGCGCCCGACGACCGGGCGGA.....GG 31135  
650 nMetProAlaProProHisProArgThrProLeuGlySerProAla... 665  
:|||||:|||||:|  
31136 TCGCCCTCGGCTGGACGCGCTGCGACGCGTGGCGCGCTGCACGAC 31185

665 ..... 665  
31186 GCGCGCCCGCGCGACGACGACATCACGACGTGTGACGACGCGGA 31235  
666 .....AlaTyrTrpLysArgValGlyHisSerAlaCy 676  
:|||||:|||||:|  
31236 CGACGACAGATCTTCGGTTCTCGAGCGCGGTACGCGACGTCTGTGAC 31285  
676 sser.....AlaSerCysGlyLysGlyVal..TrpArg 686  
:|||||:|||||:|  
31286 CCGCGCCCGGAGCCCGCGCTCATCGAAAGAACACCACTGGCGA 31335  
686 ..... 686  
31336 ACAGAGAGAGCTGCGCGCTTACCTCAAGCGGTGACGGGTGAGCTGCAC 31385  
687 ....ProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAspG 702  
:|||||:|||||:|  
31386 CCGGCCACCGACGCTCGTGGCTGGACCGGGCGGCCACGAGCCGAT 31435  
702 luArgSerCysAlaAlaGlyAlaArgProProAlaSerProGlu.....716  
:|||||:|||||:|  
31436 CGCGATCTCGGCGGCTGCGGCTGCGGCTCCCGGGCGGTGAGAGTCCGG 31485  
717 ...ProCysHisGlyThrProCys.....723  
:|||||:|||||:|  
31486 AGGACCTGT...GGAGCTGTGACGCGGTGCGGACGCGGTGCGGCCG 31532  
724 ProProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCys. 739  
:|||||:|||||:|  
31533 GCGCCCGCGGACCGCGGTGGGACGTGGAGGGAAGGTACTCGCCGACCC 31582  
740 .....GlyProGlyThrGlnHisArgGlnLeuGlnCysArgGln 752  
:|||||:|||||:|  
31583 CGACACGCGCGCACCTCGTACTGCGCGGAGGGCGGTCTGTGAGG... 31629  
753 GluPheGlyGlyGlySerSerValProGluArgCysGlyHisLe 769  
:|||||:|||||:|  
31630 .....GGCGGACCGGTTCGACCCGCGCTCTTCGCGATCTCGGCC 31670  
769 uproArgProasn.....IleThrGlnSerCysGlnLeuArgLeu 783  
:|||||:|||||:|  
31671 AACGAGCGCTCACCATGACCCCGAGCGGCTGTGCTGGAGACCTC 31720  
783 ysGly...HisTrpGlu.....ValGlySerProTrpSerGln 794  
:|||||:|||||:|  
31721 CTGGGAGCGCTGGAGCGCGGTCTGGACCCCGCTCCCTGGCGGCA 31770  
795 CysSerValArgCys.....GlyArgGlyGlnArgSerArgGlnValar 809  
:|||||:|||||:|  
31771 GCGGACCGCGGTGTTCGCGGGCGGTGGAGCGGCTACCGAGAGGCG 31820  
809 gCysValGlyAsn.....AsnGlyAspGluValSerG 820  
:|||||:|||||:|  
31821 GTCGAAGGCTCGAAGCGCATCTGAGCGCCCACTCTGCGCGGCTCGT 31870  
820 luGlnGluCysAlaSerGlyProProGlnPro.....ProSerArg 833  
:|||||:|||||:|  
31871 CAGCTTCACCGCGCGCTGCGCTACGCGCTGGGCGCTGGAGGCGCGG 31920  
834 GluAlaCysAspMetGlyProCysThrThrAlaTrpPheHisSerAspTr 850  
:|||||:|||||:|  
31921 CGCTACGATCGACACGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTG 31970  
850 p..... 850  
31971 GCGGTGACGTCACCTGCGCGGGGCGAGTGCAGACCTCGCACTGCGCGCGG 32020  
851 .....SerSerLysValSerProGluProPro 859  
:|||||:|||||:|  
32021 CCGCAGGTCATCGCGGACTTCGCGCTCTTACCCAGTTCTCCC 32064

Align seg 1/1 to reverse of: US-07-945-283-1 from: 1 to: 8438

alignment\_block:  
US-10-041-770-2 x US-07-945-283-1/rev

```
307 .....LysGlyGluSerLeuSerAlaGluGlyPro...ThrThrG 578
5603 CCTCGTCTCTCCGAGGAGAGAGAGAGAGAGTCCGGCCCGCGGC 5554
308 .....GlyArgGlyGln...GlnGlyGlnGlyProTyrP 318
5553 CCACCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5504
318 lyThrGlyGlyThrProHisGlyProArgLeuGluProAspP 334
5503 CCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5454
335 ProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisSer 351
5453 CCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5404
351 u.....TrpSerLeuPheAlaProSerS 359
5403 CGCTCCACCAACACCACTCTCGCTCATGCGCGAGC...GGCCCGCGC 5357
359 erProIleProArgCysSerGlyGluSerGlnGlnLeuArgAlaCysSer 375
5356 GACCGAGCGCGCGCTGCTCACCC.....CGCTCGGGGAGC 5322
376 GlnAlaProCysProGluGlnProAspProArgAlaLeuGlnCysAl 392
5321 CCTGGCGCGCTCCGACCTCCGCGCGCGCGCGCGCGCTCCGCTACGGCGG 5272
392 aAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrGlnTrpGluProP 409
5271 G.....CCGG 5267
409 heTrpGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProArg 425
5266 GGACTCCCGCGAGGGCTCTGGGACGAGGAGGAGCTCCGCGCAGCGCGG 5217
426 GlyPheArgPheTyrValArgHisThrGluLysValGlnAspGlyThrLe 442
5216 .....CCCGCTACCGC 5206
442 ucysGlnProGlyAlaProAspIleCys..... 451
5205 GCGCGCGCGCGCGCG.....TGCCGCTCTTCATCCCGGAGATGGG 5165
452 ..ValAlaGlyArgCysLeuSerProGlyCys...AspGlyIleLeuGly 466
5164 GGACTCGAGAGAGACGACGACGAGCGCTCGTGCCTCATCTACAGCGGG 5115
467 SerGlyArgArgPro...AspGlyCysGlyValCysGlyGlyAspAspSe 482
5114 CGCGCGCGAGGCGCATGCTCTGGCTGCAGAACCCCGCGGA..... 5076
482 rThrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuG 499
5075 ....TGCAGGCGCGCGACGAGCTTCAACACAGTTCGCCAGC..... 5037
499 lyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLeuArgLeuGlnIle 515
5036 .....GCCGGGTCCACGCGC.....CCC 5019
516 AlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyG 532
5018 ACGGCGACGGCTCTTCATCACCG..... 4995
532 yArgSerIleLeuAsnGlyAsnTrpAlaVal..AspProProGlySerTyr 548
4994 .....GCAGCGTGACCGCGCGCGCTGCCGCA 4970
549 ArgAlaGlyGlyThrValPheArgTyrAsnArgProProArgGluGluG 565
4969 CATCGGGGA.....CGCATGCGCGCGCGCGCGCTCTGGG 4932
```

4236 AGCACAAGTCGGCTCCAGCAGCAGTCCAGTCCGGCA..... 4197  
848 eAspTrpSerSerIysValSerProGluProAla 860  
4196 .....CCAAAGTCGGCTCCGGCTCCCTCCGGCT 4167  
seq\_name: /cgn2\_5/ptodata/2/ina/6B\_COMB.seq:US-09-103-840A-1

seq\_documentation\_block:  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

alignment\_scores:  
Quality: 359.50 Length: 938  
Ratio: 0.999 Gaps: 50  
Percent Similarity: 38.380 Percent Identity: 25.160

alignment\_block:

US-10-041-770-2 x US-09-103-840A-1/rev ..  
Align seq 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529  
74 ProThrValGlnLeuHisProSerLeuProAlaProAlaProAla 90  
9394104CCGGCGCGCGCTGGCGCGCGCGCTGGCGCGCTGGCGCGCTG 3934055  
90 gHisProGluAlaLeuProAlaProAlaProAlaProAla 107  
9394054AAAGCGGAGAGCCAGCGCGAGACCGCTGGCGCGCGCTGCG 3934005  
107 exProGlu.....ThrLeuProLeuTyr.Argth 116  
9394004CGCGCGCGCGCTTCCGCGCGCGCGCTTACCGCGCGCGCGCG 3933955  
116 rGlnSerArgGlyGlyGlyProLeuArgGlyProAlaSerHisLeuG 133  
9393954AAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933924  
133 lYArgGluGluThrGlnGluLeuArgAlaAlaArgSerArgLeuArg 149  
9393923.....CGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933888  
150 AspProLeuLysProGlyMetPheGlyTyrGlyArgValProPheAla 166  
9393887.....CCAGGG.....TT 3933880  
166 uProLeuHisArgAsnArg.....ArgHis..ProArgSerProAla 180  
9393879CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933830  
180 gSerGluLeuSerLeuLeuSerSerArgGlyGluGluAlaLeuProSerP 197  
9393829GCGCGCGCGCTTGCGCTTGATC.....GAGGTGGTCTTGCGCGCG 3933792

197 roThrProArgAlaGluProPheSerAlaAsnGlySerProGlnThrGlu 213  
3933791CCAGCGCGCGCTTGGCGCGCTGG.....CGCGCGCGCGCGCG 3933754  
214 LeuProProThrGlu...LeuSerValHisThrProSerProGlnAlaG 229  
3933753TGGCGCGCGCTTGAATTTGTCGCCGCCCGCGCGCGCGCGCG 3933704  
229 uProLeuSerPro...GluThrAlaGlnThrGluValAlaProArgThrA 245  
3933703CCCGTTGCCCGCATTCGCCGGCGCGCGCGCTGCCGCGCTTGGCTG 3933654  
245 rgProAlaPro...LeuArgHisHisProArgAlaGlnAlaSerGlyThr 260  
3933653GGCGCGCGCGCGCTTGGCGCGCTTGGCGCGCGCGCGCGCGCG 3933604  
261 GluProProSerProThrHisSerLeuGlyGluGlyGlyPhePheArg 277  
3933603TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933584  
277 aserProGlnProArgArgProSerSerGlnGlyTyrAlaSerProGlnV 294  
3933583GTCCCGGAGACCTGCGGACCGGTTGCGCGCGCGCGCGCGCG 3933534  
294 aAlaGlyArgArgProAspPro...PheProSerValProArgGlyArg 309  
3933533TGC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933484  
310 GlyGlnGlnGlyGlnGlyProTyrGlyThrGlyGlyThrProHisGlyPr 326  
3933483CGCGCGCGCGCGCTTGGCGCGCTTGG.....CGCGCGTCAAAGCG 3933437  
326 oArgLeuGluPro.....AspProGlnHisProG 336  
3933436CAGCGCGGAGACCGCTGGCGCGCGCGCGCGCGCGCGCGCGCG 3933387  
336 lValaTyrLeuProLeuLeuSerAsnGlyProHisAlaSerSerLeuTyr 352  
3933386CGCGCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933349  
353 SerLeuPheAlaProSerSerProLeuProArgCysSerGlyGluSerGl 369  
3933348.....CGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933308  
369 uGlnLeuArgAlaCysSerGlnAlaProCysProProGluGln..... 383  
3933307CG 3933258  
384 .....ProAspPro.ArgAlaLeuGlnCysAlaAlaPh 394  
3933257CGGTGGGTGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933208  
394 easnSerGlnGluPheMetGlyGlnLeuTyrGlnTyrGluProPheThrG 411  
3933207C.....CGCCTGCACCGCGCGCGCGCGCGCGCGCGCGCG 3933179  
411 luValGlnGlySerGlnArgCysGluLeuAsnCysArgPro..... 424  
3933178GTGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933138  
425 .....ArgGlyPheArg..... 428  
3933137CTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933088  
429 .....PheTyrValArgHis..ThrGluLysValGlnAsp 439  
3933087GGCG 3933038  
440 GlyThrLeuCysGlnProGlyAlaProAspPheCys.....ValAlaG 454  
3933037GCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3932988



[illegible]

Percent Similarity: 40.525 Percent Identity: 24.859  
alignment\_block:  
US-10-041-770-2 x US-09-105-537-1 ..  
Align seg 1/1 to: US-09-105-537-1 from: 1 to: 15872  
28 LeuSerGlyHisSerLeuGlnThrProThrGluGlyGlnGlyProG 44  
||||:||||: : : : : : : : : : : : : : : : : : :  
4249 CTGGAGGCGCGGGCGTCTGAGTCTCTACCGAGGAGGGAGTGGCGG 4298  
44 lu.....GlyValTrpGly 48  
4299 AGACGGCGCGCGCGCGCGCGGCGATCTCTCTCGGCATCAGCGGC 4348  
49 ProTrp...ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyV 64  
||| : : : : : : : : : : : : : : : : : :  
4349 ACCAACGCCACATCTGTGTGAACAGGCCCGGAAGCGCGG..... 4391  
64 alGlnArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro 80  
||| : : : : : : : : : : : : : : : : : :  
4392 .....AGCGCGGTCACCAACCCCGCGGGAAGCAGGGAGCGG 4433  
81 SerLeuProLeuProArgProArgHis..... 91  
4434 GGGAGCGCGGACACCGCCACACACGACGCGCGCGCGTGGCGTC 4483  
92 .....ProGluAlaLeuLeuProArgGlyGln..... 100  
4484 CCGAACCCGTACGCGCCCGCGTCTCGCGCGGACGCGCGCGC 4533  
101 .....GlyProArgPro.....GlnThrSerProGlu 109  
: : : : : : : : : : : : : : : : : :  
4534 CTGCGCGCCAGCGCGTTCGGTTCGGACCTTCCTCGACGCGCGACCG 4583  
110 ThrLeuProLeuTyrArgThrGlnSerArg.GlyArgGlyGlyProLeuA 126  
||| ||| : : : : : : : : : : : : : : : : : :  
4584 ACGTCAACGTCGCGGACCTCGGAGCGTCTGCGCGCGCTACCGCTTC 4633  
126 rgGlyProAlaSerHisLeuGlyArgGluGlnThrGlnGluLeuArgAla 142  
4634 GAGCAGAGCCGCTCCACCCAGCCACGACGAGGAGAGTCTCGCGG 4683  
143 AlaArgSerArgLeuArgAspProIleLys...ProGly.....Me 156  
||||||| ||| ||| : : : : : : : : : : : : : : : : : :  
4684 GCTCGAGCCCTCGCGCGCGGGAGCAAGCACCGGCGCTGTTCACGCGG 4733  
156 tPheGlyTyrGlyArgValProPheAlaLeuProLeuHisArg..... 170  
4734 AACCGCGCGCGCGGACG...CACGCGCTCTCTGTTCACCGCGCGGGA 4780  
171 .....AsnArgArgHisProArgSerProArgSerGluLeuSerLeu 185  
||| ||| : : : : : : : : : : : : : : : : : :  
4781 GCGCAGCGCTGCCATGGCGAGGAAGTGGCGC...CGCGCACCCCGT 4827  
186 IleSerSerArgGlyGluGluAlaIleProSerProThrProArgAlaG 202  
: : : : : : : : : : : : : : : : : :  
4828 GTTCGCGCGCGCTCGACACCTGTACCGGCGCTCGACCGCTCACCTCG 4877  
202 uProPheSerAla..AsnGlySerProGlnThrGluLeuProThrGlu 218  
||| : : : : : : : : : : : : : : : : : :  
4878 ACCGCGCGTGGGAGATGTGCGCGCGGGGAGG.....AGCTG 4918  
218 uLeuSerValHisThrProSerProGlnAlaGluPro..... 230  
||| : : : : : : : : : : : : : : : : : :  
4919 GACCTCACCGCGTACACCCAGCCGCGCTCTCGCTTCGAGTGGCGGT 4968  
231 .....LeuSerProGlu 234  
4969 GTTCGCGCTCTCGAACACACCGCTCGTCCCGACCTCTCACCGGCC 5018  
235 ThrAlaGlnThrGluValAlaProArgThrArgPro..AlaProLeuArgH 251  
||| : : : : : : : : : : : : : : : : : :  
5019 ACTCGTGGCGAGATGCCGCCCGCACGTCGCCGGTGTCTCTC.... 5064  
251 isHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHis 267  
||||| : : : : : : : : : : : : : : : : : :  
5065 .....CTTCGAGCGCGCGAGTCTGTCTACCGCGCGCGCGCTCAT 5109  
268 SerLeuGlyGluGlyGlyPhePheAlaSerProGlnProArg...Ar 283  
: : : : : : : : : : : : : : : : : :  
5110 GCAFTGCGCGCGCGAGGCG..GCGGATGATCGCGCTCGAGCGGGGAG 5158  
283 gProSerSerGlnGly..... 288  
5159 GCCGAGTCTCGAGTCCCTCAAGGGTACAGGGCAGGGTCCGCTCGC 5208  
289 .....TrpAlaSerProGlnValAlaGlyArg 297  
5209 CGCGGTCAACGAGCCACCGCGTGTCTCTCG.....GCGACGCGG 5252  
298 ArgProAspProPheProSerValProArgGlyArgGlyGlnGlnGlyG 314  
||||| : : : : : : : : : : : : : : : : : :  
5253 ACGCCG...CCGAGGATCGCGCGGTATGGCGGAGCGCGCGCGCGC 5299  
314 nglyProTrpGlyThrGlyThrProHisGlyProArgLeuGluProA 331  
: : : : : : : : : : : : : : : : : :  
5300 ACCGCGAGGTGCGGTGAGCGACGCTTCCACTCCCGC..... 5339  
331 spProGlnHisProGlyAlaTrpLeu.....ProLeuLeuSerAsn 344  
5340 .....ACATGGAGCGTCTCTCGACGAGTCTCTCGG 5372  
345 GlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSerProI 361  
: : : : : : : : : : : : : : : : : :  
5373 TCGCGAGGCGCTGACCTTCGAGGAGCGCGGATCCCGTCTGTCTCCA.. 5420  
361 eProArgCysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaP 378  
||||| : : : : : : : : : : : : : : : : : :  
5421 .....CGTCAACGCGCGCTCGT 5439  
378 roCysProProGluGlnProAspProArgAlaLeuGln..... 390  
5440 CACGTCGCGGAGCTCACCTCGCGCGTACTGGTGCACGACAGATCGCG 5489  
391 .....CysAlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrG 405  
||||| : : : : : : : : : : : : : : : : : :  
5490 GCGCGTCTCTGACGCGCTCGCA.....CC 5521  
405 nTrpGluProPheThrGluValGlnGlySerGln..ArgCysGluLeuAsn 421  
||| ||| : : : : : : : : : : : : : : : : : :  
5522 CTGGCGCGCGGAGCGCGCTCTCTCGAGATCGCGCGCGCGC... 5568  
422 CysArgProArgGlyPheArg.....PheTyrValArgHisThrG 435  
||||| : : : : : : : : : : : : : : : : : :  
5569 ...CGTCTCACGCGACTCGCGGAGGAGTCTCGCGCGCGCGCGGAG 5615  
435 uLysValGlnAspGlyThrLeuCysGlnPro.....GlyAlaP 448  
: : : : : : : : : : : : : : : : : :  
5616 CCGCGAGCGCGGAGCTACGCTCGCTCGCTCGCGCGCGCGCGC 5665  
448 ro.....AspIleCysValAlaGly 454  
||| : : : : : : : : : : : : : : : : : :  
5666 CCGCGAGCGGAGACCTCGCGCGCTCTCGGACCGCGCGCTCTCGAGG 5715  
455 ArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgPr 471  
||||| : : : : : : : : : : : : : : : : : :  
5716 CCGACCTTGGACG.....GGCTGTCTTCCGCGAGCGCGCGC 5756  
471 oAspGlyCysGlyValCysGlyArgAspSerThrCysArgLeu...V 487  
||| : : : : : : : : : : : : : : : : : :  
5757 GCACGGA.....CTGCCACGCTACGCTCTCGG 5785  
487 alSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnIle 503  
: : : : : : : : : : : : : : : : : :  
5786 CCGGAGCACTACTGTGCTACGCGCGGAGCGCTACGAGCGCGCGCACT 5835

5019 ACTCGTGGCGAGATGCCGCCCGCACGTCGCCGGTGTCTCTC.... 5064  
251 isHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHis 267  
||||| : : : : : : : : : : : : : : : : : :  
5065 .....CTTCGAGCGCGCGAGTCTGTCTACCGCGCGCGCGCTCAT 5109  
268 SerLeuGlyGluGlyGlyPhePheAlaSerProGlnProArg...Ar 283  
: : : : : : : : : : : : : : : : : :  
5110 GCAFTGCGCGCGCGAGGCG..GCGGATGATCGCGCTCGAGCGGGGAG 5158  
283 gProSerSerGlnGly..... 288  
5159 GCCGAGTCTCGAGTCCCTCAAGGGTACAGGGCAGGGTCCGCTCGC 5208  
289 .....TrpAlaSerProGlnValAlaGlyArg 297  
5209 CGCGGTCAACGAGCCACCGCGTGTCTCTCG.....GCGACGCGG 5252  
298 ArgProAspProPheProSerValProArgGlyArgGlyGlnGlnGlyG 314  
||||| : : : : : : : : : : : : : : : : : :  
5253 ACGCCG...CCGAGGATCGCGCGGTATGGCGGAGCGCGCGCGCGC 5299  
314 nglyProTrpGlyThrGlyThrProHisGlyProArgLeuGluProA 331  
: : : : : : : : : : : : : : : : : :  
5300 ACCGCGAGGTGCGGTGAGCGACGCTTCCACTCCCGC..... 5339  
331 spProGlnHisProGlyAlaTrpLeu.....ProLeuLeuSerAsn 344  
5340 .....ACATGGAGCGTCTCTCGACGAGTCTCTCGG 5372  
345 GlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSerProI 361  
: : : : : : : : : : : : : : : : : :  
5373 TCGCGAGGCGCTGACCTTCGAGGAGCGCGGATCCCGTCTGTCTCCA.. 5420  
361 eProArgCysSerGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaP 378  
||||| : : : : : : : : : : : : : : : : : :  
5421 .....CGTCAACGCGCGCTCGT 5439  
378 roCysProProGluGlnProAspProArgAlaLeuGln..... 390  
5440 CACGTCGCGGAGCTCACCTCGCGCGTACTGGTGCACGACAGATCGCG 5489  
391 .....CysAlaAlaPheAsnSerGlnGluPheMetGlyGlnLeuTyrG 405  
||||| : : : : : : : : : : : : : : : : : :  
5490 GCGCGTCTCTGACGCGCTCGCA.....CC 5521  
405 nTrpGluProPheThrGluValGlnGlySerGln..ArgCysGluLeuAsn 421  
||| ||| : : : : : : : : : : : : : : : : : :  
5522 CTGGCGCGCGGAGCGCGCTCTCTCGAGATCGCGCGCGCGC... 5568  
422 CysArgProArgGlyPheArg.....PheTyrValArgHisThrG 435  
||||| : : : : : : : : : : : : : : : : : :  
5569 ...CGTCTCACGCGACTCGCGGAGGAGTCTCGCGCGCGCGGAG 5615  
435 uLysValGlnAspGlyThrLeuCysGlnPro.....GlyAlaP 448  
: : : : : : : : : : : : : : : : : :  
5616 CCGCGAGCGCGGAGCTACGCTCGCTCGCTCGCGCGCGCGCGC 5665  
448 ro.....AspIleCysValAlaGly 454  
||| : : : : : : : : : : : : : : : : : :  
5666 CCGCGAGCGGAGACCTCGCGCGCTCTCGGACCGCGCGCTCTCGAGG 5715  
455 ArgCysLeuSerProGlyCysAspGlyIleLeuGlySerGlyArgArgPr 471  
||||| : : : : : : : : : : : : : : : : : :  
5716 CCGACCTTGGACG.....GGCTGTCTTCCGCGAGCGCGCGC 5756  
471 oAspGlyCysGlyValCysGlyArgAspSerThrCysArgLeu...V 487  
||| : : : : : : : : : : : : : : : : : :  
5757 GCACGGA.....CTGCCACGCTACGCTCTCGG 5785  
487 alSerGlyAsnLeuThrAspArgGlyGlyProLeuGlyTyrGlnIle 503  
: : : : : : : : : : : : : : : : : :  
5786 CCGGAGCACTACTGTGCTACGCGCGGAGCGCTACGAGCGCGCGCACT 5835

seq\_name: /cgn2\_6/ptodata/2/1na/6B\_COMB.seq:US-08-506-296B-3

```

NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSRI 489.0
TELEPHONE: (619) 554-2937
TELEFAX: (619) 554-6312
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3991 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 59...3859
US-08-506-296B-3

alignment_scores:
Quality: 347.00 Length: 1103
Ratio: 0.801 Gaps: 64
Percent Similarity: 39.257 Percent Identity: 24.257

alignment_block:
US-10-041-770-2 x US-08-506-296B-3 ..

Align seg 1/1 to: US-08-506-296B-3 from: 1 to: 3991

27 ValLeuSerGlyHisSerLeuGlnThrProThrGlugluGlyGlnGlyPr 43
|||||
27 GTGAGACGAGCGCGCAGCTGCCCGTCCGC
|||
43 oGluGlyValTrpGlyProTrpVal...GlnTrpAlaSerCysSerGlnP 59
|||||
57 .....CCATGGCTCTGCCCATGTGTGGCTCTCTCTCTGCT 90
|||||
59 roCysGlyValGlyValGlnArg..... 56
|||
91 CCTGTCTGGGGGGCGCGAGCGCCATCACATTCCC CGGAGTAGTG 140
|||||
67 ...ArgSerArgThrCysGlnLeuProThrValGlnLeuHisPro..... 80
|||||
141 GTGGCAGCAATTTCCTGCAGCCCCCGAGCTGACGAGGAACCCCGCGAA 190
|||||
81 .....SerLeuProLeu.....Prof 86
|||
191 CAACTCGTGGTCTCCCGAGTAGATGATCGTCTCAAATGGGTGGCCAC 240
|||||
86 ro.ArgProProArg.....HisPro.....GluAlaLeuLe 96
|||
241 CGGAACCCCGCTCCAGTAGCCGATGGAGCCGTGAGGATCAGCCCTTCG 290
|||||
96 uProArgGlyGln.....GlyProArg..... 103
|||||
291 TCCCGGAGGACACGGGGGTCTCGGTGTGTCCCGGATCGGGGACTTTG 340
|||||
104 .....ProGlnThrSerProGlnThrLeuProLeuTyYr 114
|||||
341 GTCATCAACGCCACGTTGGCGCGGCTCCAGGGGCGCTTCGCGTGTCT 390
|||||
115 ArgThrGln.....SerArgGlyArg..... 121
|||||
391 CGCCACCAACGCGTTGGCACC GGCTGTCTCCCGAGGCCAAGCTCATGG 440
|||||
121 ..... 121

441 CCGAGAACAATCCGAGTGGCGCGAAGGAGAGGTGACCCCGTGGAGGTG 490
|||||
122 ....GlyGlyProLeuArgGlyProAlaSerHisLeuGlyArgGluGlu 136
|||||

```



[illegible]

21103 CGGCGAGACCCGCTGCTGGCGCGCGGTGGCGCTCGGGACTCGGAC 21152  
||||| :||| :|||  
373 ...AlaCysSerGlnAlaProCysProGluGlnProAsp..... 385  
||||| :||| :|||  
21153 GGCTGCCTGCTCACGGGAGGCTCTCCCTCGTAGCACCCTCGGTGGC 21202  
||||| :||| :|||  
386 .....ProArgAlaLeuGlnCysAlaAlaPheAenSerGln 397  
||||| :||| :|||  
21203 GGAACACGCGTGGCGGACACCGTGCTGCTCGCGGAGACGCGCTCG... 21249  
||||| :||| :|||  
398 GluPheMetGlyGlnLeuTyrrGlnTrp.....GluProPheThrGlu.. 411  
:||| :||| :|||  
21250 .....TGAGCTGGGTTCGAGCGCGGGGCCAGCGTC 21281  
||||| :||| :|||  
412 .....ValGlnIserGlnArgCysGluLeuAsnCysArgP 424  
:||| :||| :|||  
21282 GGTTCGATCTGGTCGAGGAGCTCACTCGACGCGCGGTGCTGCTGC 21331  
||||| :||| :|||  
424 roArgGlyPheArgPheTyrrValargHisThrGluLysValGlnAspGly 440  
:||| :||| :|||  
21332 CC.....TGCTGGCGCGG 21345  
||||| :||| :|||  
441 ThrLeuCysGlnProGlyAlaProAspileCysValAlaglyArgCysLe 457  
:||| :||| :|||  
21346 TCCGNGTCCAG.....TGCTCGTCGGCGCGAGCGAC 21377  
||||| :||| :|||  
457 userProglyCysaspGlylle.LeuglySerGlyArgArgProAspGly 473  
:||| :||| :|||  
21378 GAGTCCGGCGCTC.....GTACTTCGGGCTCTACGCGCACCCCGGAGA 21421  
:||| :||| :|||  
474 CysGlyValCysGlyYAspAspSerThrCysArgLeuValSerGlyAs 490  
:||| :||| :|||  
21422 CGCGCGGCGGCGGAGTGGACGCGCACGCCCGGTGTGCTGGCG 21471  
||||| :||| :|||  
490 nLeuthrAspArgGlyProLeuglyTyrrGlnLysIleLeuTrpIleP 507  
:||| :||| :|||  
21472 C.....CGGTGGGACCGACCGCCCCGCTCGCGGAC 21503  
||| :||| :|||  
507 roAlagly..... 509  
:||| :||| :|||  
21504 CCGGAGGCTGTCGCCCGCGCGCGCGGTTGGACGTGGAGGTCT 21553  
||||| :||| :|||  
510 .....AlaLeuArgLeuGlnIleAlaGlnLeuArgProSer..... 521  
||||| :||| :|||  
21554 GTACGAGCGCTTCGGCGGACGCGTACGGTACGGCCCTCTCCAGG 21603  
||||| :||| :|||  
522 .....SerAsnTyrrLeuAlaLeuArgGlyP 530  
:||| :||| :|||  
21604 GCCTCCGTGTCTCGCGCGGTGGCGAGGTGTTCGCGAGCTGGGCC 21653  
||| :||| :|||  
530 roGlyArgSerIleIleAsnGlyAsnTrpAlaValAspPro...Pro 545  
:||| :||| :|||  
21654 CTCGGCGCGAGGTCCGGTCCGAGGCGCGGTTCCGCCITCACCC 21703  
||| :||| :|||  
546 GlySer.....TyrrAglaGlyThrVal..... 554  
:||| :||| :|||  
21704 GCGCTGCTCGACGCGCGTCCAGCGCGCGGTGCGGCGGGCGGTTCG 21753  
||| :||| :|||  
555 .....P 555  
||| :||| :|||  
21754 GCGCGGGCACGGGTGCGGTTCGCTGGAGCGGATCTCCCTGTACGCG 21803  
||| :||| :|||  
555 heArgTyrrAsnArgProArgGluGluGly..... 565  
||| :||| :|||  
21804 GTCGGCGCCACGCGCTCGCGTGGGGTGGCCCCCGCGCGCGGACAC 21853  
||| :||| :|||  
566 .....LysGlyGluSerLeuSerAlaGluGly..... 574  
:||| :||| :|||  
21854 GGTGTCGTGACGCGCGCGCATCTCCGGGACGCGGTGTTCGCGCGGG 21903  
||| :||| :|||  
575 ....Pro,ThrThrinProValAspValTyrrMetIlePheGlnGlu 589  
:||| :||| :|||

[illegible]

[illegible]

```

813 nAsnGlyAspGluValSerGluGlnGluCysAlaSerGlyProProGlnP 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
22774 .....GACAGTCCCGATCGCGATCCGGCCACCGGTCT 22807
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
830 roProSerArgGluAlaCysAspMetGlyProCysThrThraL..... 844
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
22808 CAACTTCGGGACGTCTGTATCGCCCTCGGCATGTACCCCGATCCGGCGC 22857
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
845 ...TrpPheHisSerAsp.....TrpSerSerLysValSerPro..... 856
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
22858 TGAATGGCACCGAGGAGCGGGGTGTGTACCGCGCCCGCCGGCGTC 22907
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
857 .....GluProProAla 860
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
22908 ACGCACCTCGCCCGCGCGC 22926

seq_name: /cgn2_5/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

alignment_scores:
  Quality: 344.50      Length: 957
  Ratio: 0.946        Gaps: 54
  Percent Similarity: 38.036  Percent Identity: 25.392

alignment_block:
US-10-041-770-2 x US-09-103-840A-2/rev  ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

74  ProthrValGlnLeuHisProSerLeuProLeuProAArgProProAr 90
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
3927959CGGCGCGCGGCTTCCGCGCGCCGCCCTTGGCCGCTTGGCCGCGCTC 3927910

90  gHisProGluAlaLeuLeuProArgGlyGlnGlyProArgProGluThrS 107
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
3927909AAAGCCGGAGGAGCCGACGCGGAGACCGCTGGCTCCTCGCCG...CCGT 3927863

107  srProGluThrLeuPro..LeuTyArGThrGlnSerArgGlyArgGlyG 123
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
3927862CACCGCCCAACACCGCGTTTCCGCGCGCGCGCGCTTACCGCGGG.... 3927818

123  yProLeuArgGlyProAlaSerHisLeuGlyArgGluGluThrGlnGluI 140
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
3927817..CCCGCGCGAAGCGCGCGCCACAC...CAGCGCGCGCAFCGCGCGCGCTC 3927772

140  lEArgAlaAlaArgSerArgLeuArgAspProIleLysProGlyMet 156
      |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

```



425 ...ArgGlyPheArg.....PheT 430  
 3926580AACCCGGCGCACCGGTAGCGTGTCCAGCGCGCGACCGACCTGC 39265931  
 430 yrValargHis.ThrGluIysValGlnAspGlyThrLeuCysGlnProGlu 446  
 39265930ACCACGACGACACACACACCGGTACCGCGGTACCGCGATCCCCGG 39265881  
 446 yAlaProAspIleCys.....ValAlaGlyArgCysLeuSerProGlyC 461  
 39265880CGTTACGAGCGTCCACCGCGCGCGGTGCGCGGTACCGATGGCG 39265831  
 461 ysAspGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyValCys 477  
 39265830CCACCGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 39265781  
 478 GlyGlyAspAsp..... 481  
 39265780GGCGCGCGCGATCGCGTGGGTATCGCGCGCGCGCGCGCGCGCGCC 39265731  
 482 .....SerThrCysArgLeuVal..... 487  
 39265730GGCGCGCGCTTTCGCGCGGTGCGCGGTAGCGGTGCGGTGCGCGCGCG 39265681  
 487 ..... 487  
 39265680CCGCGCGCGCGTGTACCGCTTCCGCGCGCGCGCGCGCGCGCGCAC 39265631  
 488 .....SerGlyAsnLeu..... 491  
 39265630GGCGGACCGCGTGCCTGCGCGGTGCGCGGTGCGCGGTGCGCGCGCG 39265581  
 492 ..... 492  
 39265580CGCCGCGCGCGCGCGCGCTTTCGCGCGGTGCGCGGTGCGCGGTGCG 39265531  
 492 hrAspArgGlyGlyProLeuGlyTyrClnLysIle..... 503  
 39265530CCGATGCGGTGGGTGTCGCGCGCGACACCGCGCGACACCGCGCGCG 39265481  
 504 .....LeuTyrIleProAlaGlyAlaLeuArgLeuGlnIle 515  
 39265480CGCGCGCGCGTGCCTTCCGCGCGCGCGCGT..... 39265442  
 515 eAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuArgGlyProGlyG 532  
 39265441.....GTGCGCGCACCGCGGTGTCGCGCGCGCGCGCGCGCGCG 39265405  
 532 lyArgSerIleAspGlyAsnTyrAlaValAspProProGlySerTyr 548  
 39265404CGCGCGCGGTCCACCGGAAA.....CCGCGTGC... 39265373  
 549 ArgAlaGlyGlyThrValPheArgTyrAsnArgProProArgGluGlu 555  
 39265372...GCCGTGCGACCGGTAGCGGTGCGATCCGTGCGCGCGCTGCCCGCG 39265326  
 555 ylyGlyGluSerLeuSer.....AlaGluG 574  
 39265325CGCGCGCGCGTTCGCGCGGTGCGCGGTATCCGCGCTTCCCGCG 39265276  
 574 ly.ProThrThrGlnProValPheArgValTyrMetIlePheGlnGluAs 590  
 39265276CGCGCGACACCGCGCGGT..... 39265255  
 590 nProGlyValPheTyrGlnTyrValIleSerSerProProProIleLeuG 607  
 39265254.....CCTCGGTGCGCG 3926242  
 607 luAsnProThrProGluProProValProGlnLeuGlnPro..... 620  
 3926241CGCGCGCG...CGCGCGCGCGCGCGGTTCACCGCGCGCGCGCGCGCG 3926195

```

621 .....GluIleLeuArgValGluProPr 628
3926194GGCGCGCGCGCAATCGCTGGCATATACGACCGGAGCCAC 39261445
628 oLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuGlnArgG 645
3926144AARGCGCGCGCGCGCTTGGCGCGCGTGGCGCGGACCGGAGC 3926095
645 lnValArgIleProGlnMetProAlaProHisProAlaThrProLeu 661
3926094CGTTGATGCGCGCGCAATGCGCTGGCGCGCTGGCGCGGCGCG 3926045
662 GlySerProAlaAlaTyTrpLysArgValGlyHisSerAlaCysSerAl 678
3926044GGCGCGCGCGGACCGGTAACCGGTT..... 3926015
678 aserCysGlyLysGlyValTrpArgProIlePheLeuCysIleSerArgG 695
3926014.....AGACCTGTG..... 3926006
695 luSerGlyGluGluLeuAspGluArgSerCysAlaAlaGlyAlaArgPro 711
3926005.....CTGGCGGGCGCTGGCGCGCGCTCCGCGACCG 3925970
712 ProAlaSerPro...GluProCysHisGlyThrPro.....Cy 723
3925969CAGCCACCCCTGGCGCGCGTCCGCCAACCCACGACCGCCGCG 3925920
723 sProToTyTrpGlu..AlaGlyGluTrpThrSerCysSerArgSerCys 739
3925919GCTCTGGCGCTGAAGAGCTGGCGCGGACACCGCGCGCGAGGC 3925870
740 GlyProGlyThrGlnHisArgGlnLeu...GlnCysArgGlnGluPheG 755
3925869CAGCGCGCGGACCGCGCTTACCACAGTCCACCGACAC..... 3925825
755 yGlyGlySerSerValProProGluArgCysGlyHisLeuProArgP 772
3925824.....CCCCAGAACGC.....CATCTCCGCGCGC 3925800
772 roAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisTrpGluVal 788
3925799CCGCGCGTGGCGCTGTCCGCGCGCG..... 3925771
789 GlySerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSe 805
3925770.....CGCGAGGCGCGCTGTCCGCCA 3925748
805 rArg.GlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGln 821
3925747GCGGCCAGCAGCTGGCGCGGACCGCGG..... 3925716
822 GluCysAlaSerGly.ProPro..... 828
3925715.....GCCACGCGCGCGCGGTACCGCGCTCCGCGGTACCAACCA 3925672
829 .GlnProProSerArgGluAlaCysAspMetGlyProCysThrThrAla 844
3925671ACGGCGCACCATCGTGGCGGACCAACCTGCCCGCTTGGCGCGCG 3925622
845 TrpPheHisSerAspTrpSerSerLysValSerProGluProProAlaTr 861
3925621.....CCGCGCGCGCGCGCGGTGC 3925602
861 eSer 862
3925601CAGC 3925598
seq_name: /cqn2_6/ptodata/2/lna/5B_COMB.seq:US-08-804-227C-13
seq_documentation_block:
; Sequence 13, Application US/08804227C
Patent No. 5876991

```

seq name: /cqn2 6/ptodata/2/ina/5B\_COMB.seq:US-08-804-227C-13

```
seq_documentation_block:
; Sequence 13, Application US/08804227C
; Patent No. 5876991
```

10366 .....CGGCCCATTCACGCCG 10355  
171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSerS 188  
10349 CAACCGACACCCACT. .... 10335  
188 eArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 204  
10334 .....CCGCCAACCTCAACAAGCCCC... 10311  
205 SerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerVa 221  
10310 .....GACCCACGACTCCAGCGCCCCCAGCAGCAGCATCCCC 10273  
221 lHisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
10272 CCACCCACGGCAACCCACCATCAACCCACCGCGCGCCACCAACCG 10223  
236 la.....GlnThrGlu 239  
10222 CACCGCATCCCAACGACAAACACCCCGCCACATACGACGACGCACCT 10173  
240 ValAlaProArgThrArgProAlaProLeuArgHisHisProArgAlaGl 256  
10172 CCCCACCGAATGACCAACACACCCAGCAGCATCTCCACACCGCAGCCTC 10123  
256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluGly 273  
10122 AACCGCGGAACAGCCACT...CCACGCAAAACAG... 10086  
273 lypHeArgAlaSerProGlnProArgProArgProSerSerGlnGlyTrp 289  
10085 .....CAGGCTGAGCAAACTCCGTCGACCCCAACCAACCCCAAAA 10044  
290 AlaSer.....ProGlnValAlaGlyArgArgPr 299  
10043 CCGAGTCAGATCCGCGNACATCATCCCGACAGCGGACCATCCACC 9994  
299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyP 316  
9993 TCCGCCACCGACCAACACACTCATCCAGCAGCTCCGCAACACACCCGCC 9944  
316 roTrpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspPro 332  
9943 CCGCGCATACAAACCCGCCCATACCAACCCAGCGCACCTGTAC... 9897  
333 GlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSe 349  
9896 .....CCGTGACACACCCCAACCCACCGACGACCC 9868  
349 rSerLeuTrpSerLeuPheAlaProSerSerProIleProArgGlySerG 366  
9867 C.....GCACCGCATCCCGCGCACCA..... 9846  
366 lylGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGlu 382  
9845 ..CAGCACCGGCCCGGAAGACACACCCACAGAGCCCGCGACGCCACCA 9798  
383 GlnProAspProArgAla.....GlnGlnCysAlaAlaPheAsnSerGlnGluP 399  
9797 CATCCAGACCGCGCAGCACTC.....C 9775  
399 heMetGlyGlnLeuTyrglnTrpGluProPheThrGluValGlnGlySer 415  
9774 TCACCGCCCGCGCCACACCCACCGACCGCTCAACGGAA..... 9737  
416 GlnArgCysGluLeuAsnCysArgProArg.....GlyPheArgPh 429  
9736 CGACGCGCGGACACCAACGAGCCCGCCACATCCCGCGGATCCGCAC 9687  
429 eTyrgValArgHisThrGluLysVal.GlnAspGlyThrLeuCysGlnPro 445

691	sIleSerArgGluSerGlyGluIuLeuaspGluArgSer.....	704
692	: : :	
693	-----	
694	-----	
695	-----	
696	-----	
697	-----	
698	-----	
699	-----	
700	-----	
701	-----	
702	-----	
703	-----	
704	-----	
705	-----	
706	-----	
707	-----	
708	-----	
709	-----	
710	-----	
711	-----	
712	-----	
713	-----	
714	-----	
715	-----	
716	-----	
717	-----	
718	-----	
719	-----	
720	-----	
721	-----	
722	-----	
723	-----	
724	-----	
725	-----	
726	-----	
727	-----	
728	-----	
729	-----	
730	-----	
731	-----	
732	-----	
733	-----	
734	-----	
735	-----	
736	-----	
737	-----	
738	-----	
739	-----	
740	-----	
741	-----	
742	-----	
743	-----	
744	-----	
745	-----	
746	-----	
747	-----	
748	-----	
749	-----	
750	-----	
751	-----	
752	-----	
753	-----	
754	-----	
755	-----	
756	-----	
757	-----	
758	-----	
759	-----	
760	-----	
761	-----	
762	-----	
763	-----	
764	-----	
765	-----	
766	-----	
767	-----	
768	-----	
769	-----	
770	-----	
771	-----	
772	-----	
773	-----	
774	-----	
775	-----	
776	-----	
777	-----	
778	-----	
779	-----	
780	-----	
781	-----	
782	-----	
783	-----	
784	-----	
785	-----	
786	-----	
787	-----	
788	-----	
789	-----	
790	-----	
791	-----	
792	-----	
793	-----	
794	-----	
795	-----	
796	-----	
797	-----	
798	-----	
799	-----	
800	-----	
801	-----	
802	-----	
803	-----	
804	-----	
805	-----	
806	-----	
807	-----	
808	-----	
809	-----	
810	-----	
811	-----	
812	-----	
813	-----	
814	-----	
815	-----	
816	-----	
817	-----	
818	-----	
819	-----	
820	-----	
821	-----	
822	-----	
823	-----	
824	-----	
825	-----	
826	-----	
827	-----	
828	-----	
829	-----	
830	-----	
831	-----	
832	-----	
833	-----	
834	-----	
835	-----	
836	-----	
837	-----	
838	-----	
839	-----	
840	-----	
841	-----	
842	-----	
843	-----	
844	-----	
845	-----	
846	-----	
847	-----	
848	-----	
849	-----	
850	-----	
851	-----	
852	-----	
853	-----	
854	-----	
855	-----	
856	-----	
857	-----	
858	-----	
859	-----	

```
seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
```

APPLICANT: Rosteck, Paul R., Jr.  
 APPLICANT: Sutton, Kimberly L.  
 TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: THOMAS G. PLANT 1501  
 STREET: LILLY CORPORATE CENTER  
 CITY: INDIANAPOLIS  
 STATE: IN  
 COUNTRY: USA  
 ZIP: 46285  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: ASCII(DOS) Text only  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/804,227C  
 FILING DATE: February 21, 1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Plant, Thomas, G.  
 REGISTRATION NUMBER: 35,784  
 REFERENCE/DOCKET NUMBER: X-8231  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 317-276-2459  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 44377 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 350..14002  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 14046..20036  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 20110..31284  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 31329..36071  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 36155..41830  
 US-08-804-227C-7  
 alignment\_scores:  
 Quality: 342.00 Length: 954  
 Ratio: 0.969 Gaps: 42  
 Percent Similarity: 37.002 Percent Identity: 22.432  
 alignment\_block:  
 US-10-041-770-2 x US-08-804-227C-7/rev ..  
 Align seg 1/1 to reverse of: US-08-804-227C-7 from: 1 to: 44377  
 80 ProSerLeuProLeuProArgProArgHisProGluAlaLeuLe 96  
 10628 CCGATCCCCCACTCCCGGACCCACCCACCCACCCGAC 10579  
 96 uProArgGlyGlnGlyProArgProGlnThr.....Ser 107  
 10578 ACCACACACACCCCGCGCACCCGACCGGAACTCCACGACTCCACAC 10529  
 108 ProGluThrLeuProLeuArgThrGlnSerArgGlyGlyPr 124  
 10528 CCCCAGGAACCTCCCAACACACCGGTCCTCAACAC...CGAATGAAC 10482

124 oLeuArg.....GlyProAlaSerHisLeuGlyArgGluGluThrG 138  
 10481 CATGGACACATCAACCGCGGCACTCCACCCATCCCGCATGAGGTG 10432  
 138 InGluIleArgAlaAlaArgArgSerArgLeuArgAspProIleLysPro 154  
 10431 AACACCCGACTCCAGCACACCCACATCAACCCGACACACACCCGACCG 10382  
 155 GlyMetPheGlyTyArgValProPheAlaLeuProLeuHisArgAs 171  
 10381 .....CGGCCATTACCGCGC 10365  
 171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuIleSer 188  
 10364 CAACCGACACCCACT..... 10350  
 188 eArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 204  
 10349 .....CCCCAACCCCTCAACACCCCC... 10326  
 205 SerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerVa 221  
 10325 .....GCACCAACGACTCCGACGCCCCCACCACCATCCCT 10288  
 221 lHisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
 10287 CCACCCCGGACACCCACCATCAACCCCGCGCGCCACCCACCG 10238  
 236 la.....GlnThrGlu 239  
 10237 CACCGCATCAACCAACGACACACCCCGCCACATACGACGACGACCT 10188  
 240 ValAlaProArgThrArgProAlaProLeuArgHisHisProArgAlaG 256  
 10187 CCCCACCGAATGACCAACACACGACCATCCACACCCCGGAGCTCC 10138  
 256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluGlyG 273  
 10137 AACGCCCGGAACACGCACT...CCAACGCAAAACG..... 10101  
 273 lYPhePheArgAlaSerProGlnProArgArgProSerSerGlnGlyTrp 289  
 10100 .....CAGGCTGAGCAAACTCCGTCGACCAACCAACCCCAAA 10059  
 290 AlaSer.....ProGlnValAlaGlyArgPr 299  
 10058 CCGAGTCAGCATCCGGAACATCACTCCGCAACGACCGACCATCCACC 10009  
 299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyP 316  
 10008 TCCCCACCCGACACACCTCATCCAGCACCTCCGCAAAACACCCACCC 9959  
 316 roTrpGlyThrGlyGlyThrProHisGlyProArgLeuGluProAspPro 332  
 9958 CCGCATACACACCCCGCCCATACCAACCACTGGCGACCTGAC... 9912  
 333 GlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSe 349  
 9911 .....CCGTGAACAACACCCCAACCCCGACGACCC 9883  
 349 rSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerG 366  
 9882 C.....GCACCGCACTCCCGCGACCA..... 9861  
 366 lYgluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProGlu 382  
 9860 ..CAGCACCGGCGCCCGAAGACACACCCACAGGAGCCCGCGACGCCA 9813  
 383 GlnProAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluP 399  
 9812 CATCCAGACCCCGCGCACTC.....C 9790  
 399 heMetGlyGlnLeuTyrglnTrpGluProPheThrGluValGlnGlySer 415

```

9789 TCACGCCCGCGGCCACACACCGACCGACGCTCACCGAA..... 9752
416 GlnArgCysLeuLeuAsnCysArgProArg.....GlyPheArgPh 429
9751 CGACGCGCGCACACCAACGACCGCCACATCCCGGATCCGCAC 9702
429 eTyrValArgHisThrGluLysVal..GlnAspGlyThrLeuCysGlnPro 445
9701 CCACCCGCTCCACACA.....CTCAGCAACGCGACCGCTCGCGCCGC 9658
446 GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAs 462
9657 AACGCCCC.....TCCGACCGCGCAGAAAGCACCCACGACAA 9620
462 pGly...IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysG 478
9619 CGACCCGTCACCGCTCGCGCGCTCAGTTTCAGGCGCTCCCGTGCT 9570
478 LysGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArg 494
9569 CG..... 9568
495 GlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLe 511
9567 .....CTCCGCGCTCCGCC... 9553
511 uArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuA 528
9553 ..... 9553
528 rGlyProGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspPro 544
9552 ..GGCGTTCTCCAGGACC.....ACATGGCGTGTGGTCCCG 9517
545 .....ProGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnAr 559
9516 CTCACCCCGGAAGGCC.....GACACCGCGCGCGCGCGCACCG 9479
559 g.ProProArgGluGluGlyLysGlyGlySerLeuSerAlaGluGlyPro 575
9478 CTCACCCCGCGCGCGCACGACCGGG...TCTCGGTACAGCAGCTCCACCG 9432
576 ThrThrGlnProValAspValTyrMetIlePheGlnGluGluAsnProgl 592
9431 CACCCGGAAGCCACTCCACCTTCGACGACGCGCATCCACATGCAGCGTC 9382
592 yValPheTyrGlnTyrValIleSerSerProProIleLeuGluAsnPro 609
9381 CGCGGCAACGACCATCCCGCATCCCTCGCACCATCTGTATCACACACC 9332
609 roThrProGluProProVal...ProGlnLeuGlnProGluIleLeuArg 624
9331 CACACCGCAGCGCGCTGCGCATGCCCGATGTCGACTTCAACGACCCGA 9282
625 ValGluProProLeuAlaProAlaProArg.....ProAlaArgTh 638
9281 GCCACAACGATCCCGCGCGCATCCCGCTCCCGCTACGTGGCGCACGCGCA 9232
638 rProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaPro 655
9231 CCGCGCTCGATCGATCACCGACCGCGCTCCCGTACCGTCCGCTCCAC 9182
655 roHisProArg.....ThrProLeuGly 662
9181 CACATCCACGTCCGCGGCGCACCGCCCGCATCACCGACGCGCTCACGGA 9132
663 SerProAlaAlaTyrTrpLysArgValGlyHisSerAlaCysSerAlaSe 679
9131 TCACCCCGCTGT.....GCGCGGACCACTCGCGCGCTCAGCCCA 9091
679 rCysGlyLys..... 682

```

seq\_name: /cgn2\_6/ptodata/2/ina/5b\_comb.seq:US-08-804-198-1

seq\_documentation\_block:  
; Sequence 1, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rosteck, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 31329..36071  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36155..41830  
; US-08-804-198-1

alignment\_scores:  
Quality: 342.00 Length: 954  
Ratio: 0.969 Gaps: 42  
Percent Similarity: 37.002 Percent Identity: 22.432

alignment\_block:  
US-10-041-770-2 x US-08-804-198-1/rev ..

Align seg 1/1 to reverse of: US-08-804-198-1 from: 1 to: 44377

80 ProSerLeuProLeuProArgProArgHisProGluAlaLeuLe 96  
||||: ||| |||||:||||| |||||

10628 CCGGATCCCCCAACTCCCGGAACCCACCACCCACCCGACCCGACACACCCGAC 10579  
96 uPro.ArgGlyGlnGlyProArgProGlnThr .....Ser 107  
||| :|||:||||  
10578 ACCACACACACCCGCGCCGACCCGACCGAACTCCCAAGACTCCACAAC 10529  
||| :|||:||||  
108 ProGluThrLeuProLeuTyArgThrGlnSerArgGlyArgGlyGlyPr 124  
||| ||||| :|||:||||  
10528 CCCCAGGAACTCCCAACACCCGCTCCATCAACAC...CGAATGAACCC 10482  
||| :|||:||||  
124 cLeuArg.....GlyProAlaSerHisLeuGlyArgGluGluThrG 138  
|||:||||| ||||| ||| :|||:||||  
10481 CATCGACACATCCACCCGCGGCACTCCACCCATCCCCCATCAGCGTG 10432  
|||:||||| ||||| ||| :|||:||||  
138 lngluLeuArgAlaAlaArgSerArgLeuArgAspProIleLysPro 154  
:||||| ||| :|||:||||  
10431 ACAACACCGACTCCAGCACACCCACATCAACCCGACCAACACCCGCGCG 10382  
|||:||||| ||||| ||| :|||:||||  
155 GlyMetPheGlyTyArgValProPheAlaLeuProLeuHisArgAs 171  
|||:||||| ||||| ||| :|||:||||  
10381 .....CGGCCGATTCACCGCG 10365  
|||:||||| ||||| ||| :|||:||||  
171 nArgArgHisProArgSerProProArgSerGluLeuSerLeuLeuSer 188  
:||||| ||||| ||| :|||:||||  
10364 CAACCGACACCCACT..... 10350  
|||:||||| ||||| ||| :|||:||||  
188 erArgGlyGluGluAlaIleProSerProThrProArgAlaGluProPhe 204  
|||:||||| ||||| ||| :|||:||||  
10349 .....CCCCCAACCCCTCAACACCCCCC... 10326  
|||:||||| ||||| ||| :|||:||||  
205 SerAlaAsnGlySerProGlnThrGluLeuProProThrGluLeuSerVa 221  
|||:||||| ||||| ||| :|||:||||  
10325 .....GCACCGGACTCCGAGCGCCGCCACCGACCATCCCC 10288  
|||:||||| ||||| ||| :|||:||||  
221 lHisThrProSer.....ProGlnAlaGluProLeuSerProGluThrA 236  
||| ||||| :|||:||||  
10287 CCACCCACCGGCAACCCACCATCAACCCCGCGCGCCACCAACCG 10238  
|||:||||| ||||| ||| :|||:||||  
236 la.....GlnThrGlu 239  
|||:||||| ||||| ||| :|||:||||  
10237 CACCGCATCCCCAAGACAACACCCCGCCACATACGACGACGACCT 10188  
|||:||||| ||||| ||| :|||:||||  
240 ValAlaProArgThrArgProAlaProLeuArgHisProArgAlaGl 256  
|||:||||| ||||| ||| :|||:||||  
10187 CCCCCACCGAATGACCCCAACACCCGACCTCCACACCCCGGACCTCC 10138  
|||:||||| ||||| ||| :|||:||||  
256 nAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluGlyG 273  
:|||:||||| ||||| ||| :|||:||||  
10137 AACGCCCGGACACGCGCACT...CCAACGCAACACACG..... 10101  
|||:||||| ||||| ||| :|||:||||  
273 lyPhePheArgAlaSerProGlnProArgProSerSerGlnGlyTrp 289  
:||||| ||| :|||:||||  
10100 .....CAGGCTGAGCAAACTCCGTCGACCCCAACACCCACCAAAA 10059  
|||:||||| ||||| ||| :|||:||||  
290 AlaSer.....ProGlnValAlaGlyArgArgPr 299  
|||:||||| ||||| ||| :|||:||||  
10058 CCGATTCAGCATCCCGGAACATCATCCCGACGACCGACCATCCACC 10009  
|||:||||| ||||| ||| :|||:||||  
299 oAspProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyP 316  
||| ||||| :|||:||||  
10008 TCCCCACCCGACGACACACACCTCATCCAGCACTCCGCAAAACACCCACC 9959  
|||:||||| ||||| ||| :|||:||||  
316 rotPrgGlyThrGlyThrProHisGlyProArgLeuGluProAspPro 332  
||| ||||| :|||:||||  
9958 CCCCAGTACAACCCCGCCCATACCAACCCACTGCGCACCCCTGAC... 9912  
|||:||||| ||||| ||| :|||:||||  
333 GlnHisProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSe 349  
|||:||||| ||||| ||| :|||:||||  
9911 .....CCGTGAACACACACCCCGACCCCGACGACCC 9883  
|||:||||| ||||| ||| :|||:||||  
349 rSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerG 366  
|||:||||| ||||| ||| :|||:||||  
9882 C.....GCACCGCACTCCCGCCGACCA..... 9861  
|||:||||| ||||| ||| :|||:||||

366 lyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProProGlu 382  
 9860 ..CAGCACCAGCCCGCCGAGACACACACACAGGAGCCCGCCGAGCACCA 9813  
 383 GlnProAspProArgAla..LeuGlnCysAlaAlaPheAsnSerGlnGluP 399  
 9812 CATCAGACCCGCGCAGCACTC..... 9790  
 399 heMetGlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGlySer 415  
 9789 TCAGCCCGCCGCGCCACACACACCGCCAGCTCACCGAA..... 9752  
 416 GlnArgCysGluLeuAsnCysArgProArg.....GlyPheArgPh 429  
 9751 CGAGCGCGCAGCACACACCAACAGCCCGCCACATCCCGGGATCCGCAC 9702  
 429 eTyrValArgHisThrGluLysVal..GlnAspGlyThrLeuCysGlnPro 445  
 9701 CCACCGGTCACACACA.....CTCAGCAACCGCAGCCGCTCGCGCCG 9658  
 446 GlyAlaProAspIleCysValAlaGlyArgCysLeuSerProGlyCysAs 462  
 9657 AACGCCCC.....TCGACCGCGCAGAAAGCACCCACGACAA 9620  
 462 pGly...IleLeuGlySerGlyArgArgProAspGlyCysGlyValCysG 478  
 9619 CGGACCCGCTCACCGGTCGGCGGCTCAGGTTCAGGCGCGTCCCGCGTCT 9570  
 478 lyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrAspArg 494  
 9569 CG..... 9568  
 495 GlyGlyProLeuGlyTyrGlnLysIleLeuTrpIleProAlaGlyAlaLe 511  
 9567 .....CTCCGCGCTCCGCC..... 9553  
 511 uArgLeuGlnIleAlaGlnLeuArgProSerSerAsnTyrLeuAlaLeuA 528  
 9553 ..... 9553  
 528 rgGlyProGlyGlyArgSerIleIleAsnGlyAsnTrpAlaValAspPro 544  
 9552 ..GGCGTCTCTCCAGGACC.....ACATGGCGTTGGTCCCG 9517  
 545 .....ProGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnAr 559  
 9516 CTCACCCGGAAGGCC.....GACACCGCGCGCCCGCACCCG 9479  
 559 g..ProProArgGluGluGlyLysGlyGlySerLeuSerAlaGluGlyPro 575  
 9478 CTCACCCCGCGCGCCAGCAGCGG...TCTCGGTGACAGCTCCACCG 9432  
 576 ThrThrGlnProValAspValTyrMetIlePheGlnGluAsnProG1 592  
 9431 CACCCGAGCCCACTCCACTTCGACGAGCGGCATCCACATGACAGCGTC 9382  
 592 yValPheTyrGlnTyrValIleSerSerProProIleLeuGluAsnP 609  
 9381 CGGGCAACGACCCATCGCGCATCCCTGCACACCATCTGATCACACCACC 9332  
 609 roThrProGluProVal...ProGlnLeuGlnProIleLeuArg 624  
 9331 CACACCCGCGAGCGCTCGCATGCCGATGCCGATGTCGACTTCAACAGCCGCA 9282  
 625 valGluProProLeuAlaProAlaProArg.....ProAlaArgTh 638  
 9281 GCACCAACGAGATCCGACCCGCTCCCGCGGTACGTGCGCCAGCAGCGCA 9232  
 638 rProGlyThrLeuGlnArgGlnValArgIleProGlnMetProAlaProP 655  
 9231 CCGCGCTCGATCGATCACCCAGCGCGTCCCGTACCGTCCGCGCTCCAC 9182

```
8302  ACGGCGCCGATCCCGACGATCGGATCGGTCACGTCGTGGCGCA... 8256
842  ThrAlaTrpPheHisSerAspTrpSerLysValSerProGluPr 858
8255  TGACCGCGGTGG.....GGCGGCGCGCTGCGCGCTCC 8224
858  OPro 859
8223  GCCT 8220
```

seq\_name: /sgn2\_6/ptodata/2/lna/6B\_COMB.seq:US-09-105-537-32

```
seq_documentation_block:
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-32
```

alignment\_scores:  
Quality: 341.50 Length: 1212  
Ratio: 0.759 Gaps: 64  
Percent Similarity: 37.129 Percent Identity: 22.937

alignment\_block:

US-10-041-770-2 x US-09-105-537-32 ..

Align seg 1/1 to: US-09-105-537-32 from: 1 to: 11220

```
46 ValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysGlyVa 62
6623 TTGTGGAGCGACTGGCGGTTTCAGCGGGCTGTCGTCGCGCG... 6664
62 lGlyValGlnArgSerArgThrCysGlnLeuProThrValGlnLeuH 79
6664 ..... 6664
79 isProSerLeuProLeuProPro.....ArgProProArg 90
6665 ..CGGTCAACGGGCTTACCGCCACCGGTTCGGGCGGACCCAGCA 6712
91 HisProGluAlaLeu.....LeuProArgGly.....GlnGlyPr 102
6713 TCCAGAGCTCGTCAGCGGTGTGAGCCGCGGTCGCGCAGCATC 6762
102 oArgProGlnThrSerProGlnThrLeuProLeuTyrArg..... 115
6763 ATCCCGCTGACTACGCTCCACAGCGCCACGTCGAGACCATCAGAG 6812
116 .....ThrGlnSerArgGlyArgGlyPro.....LeuArgGly 127
6813 CGAACTCGCGGACGCTCTGGCGGGTTGTCCCGCCAGACACCCAGGTCC 6862
128 ProAlaSer.HisLeuGlyArgGlnGluThrGlnGluLeuAlaAla 144
6863 CCTTCTCTCCACCTCGAAGCGCGCTGATCATCCGAACCCGCTCGAC 6912
144 rArgSerArgLeuArgAspPro.....IleLysProGly 155
```

```
6913 GCGGCTACTGTTACGCGCAACCTCGGCATCGTGTGGGCTTCGCGCGGC 6962
156 MetPheGlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnAr 172
6963 CGTGAACCTCGGCCACCGAGAGGTTTACCCACTTCTGTCGAGGTCA 7012
172 gArgHisProArg..SerPro.....ProArgSe 181
7013 GCGCCACCCCGTCTCATCCATGGCGTGTCCGAGACCGTCACCGGATC 7062
181 rGluLeuSerLeuIleSerSerArgGlyGluGluAlaIleProSerPro 198
7063 GGCACCTCCGCGGTGACACGCGGACAGCACCGCTTACACCTCCCT 7112
198 hrProArgAlaGluProPheSerAlaAsnGlySerProGlnThr..... 212
7113 CGCCGAGCGCTGGCCA.....ACGGCTCACGTCGACTGGGCT 7153
213 .....GluLeuProThrGluLeuSerValHisThrProSe 225
7154 CTCTCTCCCGACCGACCGACCCCGCATCTGCCACCTACGCTTC 7203
225 r...ProGlnAlaGluProLeuSerProGluThrAlaGlnThrGluVala 241
7204 CAGACCGAGCGCTACTGGCGCGACGCG.....ACCTCTCGC 7241
241 laProArgThrArgProAlaProLeuArgHisHisPro.....ArgAla 255
7242 CGCCGCTGACATCACCTCCGCGGTCTCGGGCGCGCGAGCACCGCTGC 7291
256 GlnAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGlu 272
7292 TCGGCGCGCGGTGGCGTCCGGGACTCCGAGCGCTGCTGCTCAGGGG 7341
272 yGlyPhePheArgAlaSerProGlnProArgArgProSer.....Ser 287
7342 A.....GCTCTCCCTCGTACGACCGCTGGTGGCGGACCA 7379
287 lngLyrAlaSerPro.GlnValAlaGlyArgArgProAspProPhePr 303
7380 CGCGGTGGCGGACCGTGTGCTGCTCGCGG..... 7409
303 oSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProTrp..... 317
7410 AACGGCGTTCGTGGAGCTGGGTTCCGAGCGCGGACCGAGTGGTTGC 7458
318 .....GlyThrGlyThrPro.....HisGlyProArgLeuGluProAsp 331
7459 GATCTGGTCGAGGAGTCACTCCCTCAGCGCGCTGCTGCTGCCCGCTG 7508
332 ProGlnHisProGlyValaTrpLeuProLeuLeuSerAsnGlyProHisAl 348
7509 TGGCGGTCCTGTCGTGTGTCGTGTCGTGCGCGGAGCGACGAGTCCGGC 7558
348 aserSerLeuTrpSerLeuPheAlaProSerSerProIleProArgCys 365
7559 GTCTGACCTTCGGGCTCTACGCGCACCGGAGGACGCGCGGCGCA...G 7605
365 exGlyGluSerGluGlnLeuArgAlaCysSerGlnAlaProCysProPro 381
7606 GCGGAGTGACGCGGACCGCACCGGTGTGTGCGCGCGCTGCGGACCG 7655
382 GluGlnProAspProArgAlaLeuGlnCysAlaAlaPhe..... 394
7656 CACCGCGCGCTCGCGGACCGGAGCGCTGCGCGCGCGCGCGCGCGAGC 7705
395 .....AsnSerGlnGluPheMetGlyGln..... 402
7706 CGGTGGACGTGACGGTCTGTACGAGCGCTTCCGCGGACCGCTACGCG 7755
403 .....LeuTyrGlnTrpGluPro 408
```



7756 TACGGCCCCCTCTTCACAGGGCGTCCGTGGTGTCTGGCGGGCTGGCACGA 7803

409 PheThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProAr 425

7806 GGTGTTCCGCCAGCTCGCCCTCGCGGCCGAGGTGCGCGGTGCCGAGGGCG 7855

425 gGlyPheArgPheTyrValargHisThrGluLysValGlnAspGlyThrL 442

7856 CGCGTTCCGGCCTTCACCCGGCGCTGCTCAGCCCGCGCGTGCAGGGCGCC 7905

442 euCysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSer 458

7906 GGTGCGGGCGGGCGGTTCG.....GCGCGGCACGCGGCTCCCGTTCG 7949

459 ProGlyCysaspGlyIleLeu.....G1 466

7950 CTTGAGGGGATCTCTGTACCGGGTGGCGGCACCGCCCTCCGGCTGC 7999

466 ySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAspAspSerT 483

8000 GGCTGGCCCCCGCGCGCCCGACACGTGTC.....CGTGAGC 8037

483 hrCysArgLeuValSerGlyAsnLeuThrAspArgGlyGlyProLeuGly 499

8038 GCGCGCACTCTCTCGGGCACCGGTGTTGCGCGGAGCTCCCTACGCT 8087

500 TyrGlnLysIle..... 503

8088 GCTGCGCGTCAACCCGGCGAGTGGCGGCTTCAGGCACCCACACTCTGC 8137

504 LeuTrpIleProAlaGlyAlaLeuArgLeuGlnIleAlaGlnLeuArgP 520

8138 ACGCGCTGCACCTCTGTGAGTGGACCGCTGGGACGGTGGCGCGCAGGCC 8187

520 rSerSerAsnTyrLeuAlaLeuArgGlyProGlyGlyArgSerIleIle 536

8188 CF.....GCCGGCGCGGTCTGTCTGGCGGCGCAGCGCGA 8222

537 AsnGly.....AsnTrpAlaValAspPro.....Pr 545

8223 CGGTCTCGCGCGCGCTGCGCGCGGTGGCACCGAGGTCTCTGCTCTCC 8272

545 oGlySerTyrArgAlaGlyGlyThrValPheArgTyrAsnArgProProA 562

8273 CGGACCTTACGGACCTGGTGGAGC.....CGTCGACCGGGCG 8310

562 rgGluGluGlyLysGlyGluSerLeuSerAla..... 572

8311 GAGACCGCGCGCGGACCGCTCTGTGGTGGCTGCCCCCGCGCGCGCC 8360

573 GluGlyPro..... 575

8361 CGATGGCGCGGAGCATGTCCGCGAGGCGCTCAGCGGTCTCGCGCTGA 8410

575 ..... 575

8411 TCGAGCGCTGGTGGCGCGACGAGCGTTACCGATGGCGCCTGGTGCTC 8460

576 .....ThrTh 577

8461 GTGACCGCGACGGGTGCGCGCGCGTTCGGGCGACGGCTGGGTCCAC 8510

577 rGlnProValaspValTyrMetIle.....PheGlnGluAsnP 591

8511 GGACAGCGCGCGCTCTGGGCGCTCGCGCGGTCTCGCGCACACGAGAGCC 8560

591 rGlyValPheTyrGlnTyrValIleSer..... 600

8561 CGGCG.....CGGTCTCTCTGCTCGACCTCGCGGGGAGCCCG 8601

601 .....SerProPro..... 604

8602 ACGCGCGCGGACCGCACCGCGCGGAGCGGCTGACGACCGGGAGCGCAC 8651

605 .....IleLeuGluAsnPro..... 609  
 8652 CCTCGCGGCACCTCTGGAGAGCGCCGCCCTCGGAGCGCCTCGGACCG 8701  
 609 ..... 609  
 8702 CCTCGGCTCGGGGAGCGCAGCTCGCCTCGGGGAGGGGCGCTCCTC 8751  
 610 .....ThrProGluProProValProGluLeuGlnP 620  
 8752 GTACCCCGCCTGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCTCGC 8801  
 620 roGlu.....IleLeuArgValGluProProLeuAlaPro 631  
 8802 CGCGCGCAGCGCCTCGCGCTCTCGCGCTGC...CGCGGCTCGGCGCC 8848  
 632 Ala..... 632  
 8849 TCTGGCGTCTGAGGCCCGGTAGGACGGCAGCTTGAGAGCTCAGCGG 8898  
 633 .....ProArgProAlaArgThrProGlyThrLeuGlnArg 645  
 8899 GCGCGCGCGGAGCGCGAGACC.....TCGCCCGGAGCGGCTCGGCGC 8942  
 645 lnValArgIleProGlnMetProAlaProPro..... 655  
 8943 GGGACAGGTCCGATCGCGATCCGGGCCACCGGTCTCACTTCGCGCAGC 8992  
 656 .....HisProArg 658  
 8993 TCCTGATCGCCTCGGCATGTACCGCATCGCGGCTGTGTGGCAGCGAG 9042  
 658 gThrProLeuGlySerPro...AlaAlaTyrTrpLysArg..... 670  
 9043 GAGCGCGCGGTGGTCACCGAGCCGGCGCGCTGACGACCTCGGCGCC 9092  
 671 ....ValGlyHisSerAlaCysSer...AlaSerCysGlyLysGlyVal 684  
 9093 CGGCGACCGGTCATCGGCGCTCTCTCGGCGCGCTACGCGCGCGCTCGC 9142  
 685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAs 701  
 9143 TGGCGG.....ACGGCGGAGCC 9159  
 701 pGluArgSerCysAlaAlaGlyAlaArgProProAlaSerProGluProc 718  
 9160 GTGCGCGCATCCCGAGGGGTGGACGTTCGCCAGGCGCCT...CCGT 9206  
 718 yHis.....GlyThrProCys..... 723  
 9207 GCGGTGGTTCCTGAGCGCGCTACGCCCTCGCGGACCTGGCGGAGC 9256  
 724 .....ProProTyrTrpGluAl 729  
 9257 TCAGGCGCGCGAGAGCGCTCTGTGCTCCACTCGCGCGCGGTG.....GC 9300  
 729 adGlyGluTrpThrSerCysSerArgSerCysGlyProGlyThrGlnHis 746  
 9301 GTGGGCATGGCCCGCTGCAGC.....TCGCCCGGACCTGGGCGCT 9341  
 746 rgGlnLeuGlnCysArgGlnGluPheGlyGly..... 757  
 9342 GGAGGTCCACGCACGGGAGGTTCACGGGAAGTGGGAGCGCCTCGCGCGC 9391  
 758 .....GlySerSerValProProGluArgCysGlyHisLe 769  
 9392 TGGCGCTGGACAGCGGCATCGCCTCTCCCGACCCCTGGACTTCGAG 9441  
 769 uProArg..... 771  
 9442 TCGGGTTCCTGTCGCGCTTCGGCGGGGCGGCGATGACGCTGACTGAA 9491

772 .....ProAsnIleThrGlnSerCysGln 779  
9492 CTGCTCGCCGCGGAGTTCTGTCGACCGCTCGCTGC.....GCCGTCTCG 9535  
780 LeuArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSe 796  
9536 GGCGGCGCGCGCGGTTCTGTTGAGATGGGAGAACCGACGCTCC.....GC 9579  
796 rValArgCysGly..... 800  
9580 GACGCGGAGCGGTCGCCGCGCGACCCCGGTGTCGGCTACCGCGCCTT 9629  
801 .....ArgGlyGlnArgSerArgGlnValArgCys..... 810  
9630 CGACCTGGCGGAGCGCGCGCGGAGCGGATCGCGGAGATGCTCGCGGAGG 9679  
811 .....ValGlyAsnAsnGlyAspGluValSerGluG 821  
9680 TCATCGCCCTCTTCGAGGACGGGGTCTCCGGACCTCGCCGCTCAGACCC 9729  
821 nGluCysAlaSerGlyProGlnPro.....ProS 832  
9730 TGGGAGCTGGCGCGCGCGCGCGCGCTTCGGACAGTCAGCGCGCG 9779  
832 erArgGluAlaCysAspMetGlyProCysThrAla..... 844  
9780 CCACAGGGCAAGGTCTCTCAGATGCGGTGCGGCGCTCGACCGCGAGG 9829  
845 .....TrpPh 846  
9830 GTACGGTCTGTCACCGCGCGCACCGGTGCGGTGGGGGCGATCGTGCGC 9879  
846 eHisSerAspTrpSerSerLysVal..... 854  
9880 CGGCACGTGTGGCGGAGTGGGCGCTACGACGCGCTCTGCTGTGAGCGG 9929  
855 .....SerProGluProAlaIleSerCysIleLeuGlyAsn 867  
9930 GCGGGGACGAGCGCGCGCGCGCGCGGAGCTGTCGACGAGCTGGAGG 9979  
868 HisAlaGlnAspThrSerAlaPhePro 876  
9980 CCCTGGAGCGGAGCTCGGTGGCGCG 10006  
seq\_name: /cgn\_6/ptodata/2/ina/6B\_COMB.seq:US-09-105-537-5

seq\_documentation\_block:  
; Sequence 5, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-5  
  
alignment\_scores:  
Quality: 341.50 Length: 1212  
Ratio: 0.759 Gaps: 64  
Percent Similarity: 37.129 Percent Identity: 22.937  
  
alignment\_block:

US-10-041-770-2 x US-09-105-537-5 ..  
Align seg 1/1 to: US-09-105-537-5 from: 1 to: 36778  
46 ValTrpGlyProTrpValGlnTrpAlaSerCysSerGlnProCysGlyVa 62  
22310 TTGTGGAGCGACTGGCGGGTTTCGACGGGCTGTCCGTGCGCG..... 22351  
62 lGlyValGlnArgArgSerArgThrCysGlnLeuProThrValGlnLeuH 79  
22351 ..... 22351  
79 lSProSerLeuProLeuProPro.....ArgProProArg 90  
22352 ..CGTCAACGGCGCTACCGCACCGTGTTCGGGCGCACCGACCCAGA 22399  
91 HisProGluAlaLeu.....LeuProArgGly.....GlnGlyPr 102  
22400 TCCAGAGCTGCTCAGCGGTGTGAGGCGCGGGGTCCGGCGCAGGATC 22449  
102 cArgProGlnThrSerProGluThrLeuProLeuTyrArg..... 115  
22450 ATCCCGCTCGACTAGCGCTCCACAGCGCGCCACGTCGAGACCATCGAG 22499  
116 .....ThrGlnSerArgGlyArgGlyPro.....LeuArgGly 127  
22500 CGAATCCGCGACGTCCTGGCGGGGTTCCTCCCGACAGACACCCAGGTCC 22549  
128 ProAlaSer.HisLeuGlyArgGluGluThrGlnGluLeuAlaAla 144  
22550 CTTCTTCTCCACCGCTCGAAGCGGCTGGATCACCGAACCGCGCTCGAC 22599  
144 rGArgSerArgLeuArgAspPro.....lLelysProGly 155  
22600 GCGGCTACTGTGTCACCGAACCTCGCCATCGTGTGGGCTTCGCCCCGCG 22649  
156 MetPheGlyTyrGlyArgValProPheAlaLeuProLeuHisArgAsnAr 172  
22650 CTTCAAAACCTTGGCCACCGAGAGGCTTCACCCACTTCGTGAGAGTCA 22699  
172 gArgHisProArg...SerPro.....ProArgSe 181  
22700 GGGCCACCGCGCTCTCTCACCATGGCGCTGCCGAGACCGCTCACCGGACTC 22749  
181 rGluLeuSerLeuIleSerArgGlyGluGluAlaIleProSerPro 198  
22750 GGCACCTTCGCGCGTGACACGGCGGACAGCACCGCTCACCATCTCCT 22799  
198 hrProArgAlaGluProPheSerAlaAsnGlySerProGlnThr..... 212  
22800 CGCGGAGCGCTGGGCA.....ACGGCGCTCACCCTCGACTGGGCT 22840  
213 .....GluLeuProProThrGluLeuSerValHisThrProSe 225  
22841 CTCTCTCCCGACGACCGACCGCTGCGCGGATCTGCCACCTACGCTTC 22890  
225 r...ProGlnAlaGluProLeuSerProGluThrAlaGlnThrGluVala 241  
22891 CAGACGAGCGCTACTTGGCGCGAGCGCG.....ACCTCTCGCG 22928  
241 laProArgThrArgProAlaProLeuArgHisHisPro.....ArgAla 255  
22929 CGCGGCTGACATCACCCTCGCGCGTCTGGGGCGCGCGACCGCTGCG 22978  
256 GlnAlaSerGlyThrGluProProSerProThrHisSerLeuGlyGluG 272  
22979 TCGGCGCGCGCTGGCGCTCGCGGACTCGACGCGTGTGCTCACGGGG 23028  
272 yGlyPhePheArgAlaSerProGlnProArgProSer.....SerG 287  
23029 A.....GCTCTCTCCCTCGTAGCGACCGCTCGCGCGGACCA 23066  
287 lnglyTrpAlaSerPro.GlnValAlaGlyArgArgProAspProPhePr 303

[illegible]

```
685 TrpArgProIlePheLeuCysIleSerArgGluSerGlyGluGluLeuAs 701
|||||
24830 TGGCGG.....ACGCGGGACC 24846
701 pGluArgSerCysAlaAlaGlyValArgProProAlaSerProGluProC 718
|||||
24847 GTCCGCGGATGCCGAGGGTGGAGCTTCGCCAGGGGGCT...CGGT 24893
718 yshis.....GlyThrProCys..... 723
|||||
24894 GCGGGGTGTTCCTACGCGCGTGTACGCCCTGGCGGACCTGGGGAGC 24943
724 .....ProProTyTrpGluAl 729
|||||
24944 TCAAGCCGCGGAGCGCCTCCGTGGTCCACTCCGCGCGCGGTG.....GC 24987
729 aglyGluTrpThrSerCysSerArgSerCysGlyProGlyThrGlnHisA 746
|||||
24988 GTGGGATGGCGCGCGCTGCAGC.....TCGCCGCGCACTGGGGCT 25028
746 rgGlnLeuGlnCysArgGlnGluPheGlyGly..... 757
|||||
25029 GGAGTCCAGCGCACGCGCGAGTACCGGAAGTGGGACGCCCTGGCGGCGC 25078
758 .....GlySerValProProGluArgCysGlyHisLe 769
|||||
25079 TCGGCCTGGACGAGCGGCACATCGCCTCTCCCGCACCTGGACTTCGAG 25128
769 uProArg..... 771
|||||
25129 TCGCGCTTCGTCGCTCCGCTTCCGCGGGGGGCGCATGGAGCTGTACTGAA 25178
772 .....ProAsnIleThrGlnSerCysGln 779
|||||
25179 CTCGCTGCCGCGGAGTCTGTCACGCTCGTGC.....GCCCTCG 25222
780 LeuArgLeuCysGlyHisTrpGluValGlySerProTrpSerGlnCysSe 796
|||||
25223 GCGCGCGCGCGGTTCGTGGAGATGGGAAGACCGACCTCC.....GC 25266
796 rValArgCysGly..... 800
|||||
25267 GACCGGAGCGGTTCGCGCGCGGACCGCGGTCTGGGTACCGCGCCTT 25315
801 .....ArgGlyGlnArgSerArgGlnValArgCys..... 810
|||||
25317 CGACCTGGCGGAGCGCGCGCGGCGGATCGCGGAGATCTCGCGGAG 25366
811 .....ValGlyAsnAsnGlyAspGluValSerGluG1 821
25367 TCATCGCCCTCTTCGAGGAGGGGTGCTCCGGCACCTGCCCGTCAAGCC 25416
821 nGluCysAlaSerGlyProGlnPro.....ProS 832
25417 TGGGACGTGGCGCGCGCGCGCGCTTCGCGCAGTCAGCCAGGCGCG 25466
832 erArgGluAlaCysAspMetGlyProCysThrAla..... 844
25467 CCACAGCGCAAGTCTCTCTCAGATCGCGTCCGGCTCGACCCGAGG 25516
845 .....TrpPh 846
25517 GTACGCTCTCTCTACCGCGGCGCGCGGTGGGTGGGGGCGATCGTGCC 25566
846 eHisSerAspTrpSerSerLysVal..... 854
25567 CGGACGTGTGGGAGAGTGGGGGTACGACGCTGTCTGTCTGTAGCCG 25616
855 .....SerProGluProProAlaIleSerCysIleLeuGlyAsn 867
25617 GCGGGGACGAGCGCGCGCGCGCGGAGCTGTGTACAGCTGGAGG 25666
```

```
868 HisAlaGlnAspThrSerAlaPhePro 876
|||||
25667 CCCTGGGAGCGGACGTCTCGTGGCGCG 25693
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11
```

```
seq_documentation_block:
; Sequence 11, Application US/08458568A
; Patent No: 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
; US-08-458-568A-11
```

```
alignment_scores:
Quality: 341.00 Length: 1023
Ratio: 0.881 Gaps: 58
Percent Similarity: 37.830 Percent Identity: 24.340
```

```
alignment_block:
```

```
US-10-041-770-2 x US-08-458-568A-11/rev ..
```

```
Align seg 1/1 to reverse of: US-08-458-568A-11 from: 1 to: 12001
```

```
74 ProThrValGlnLeuHisProSerLeuProLeu...ProProArgProPr 89
|||||
5859 CCACAGCGCGGCTGTCTCGCGCGCGCGCGCGCGCGCGCGGAGAGC 5810
89 oArgHis.....ProGluAlaLeu.LeuPro...ArgGlyGln 100
|||||
```

[illegible]

364 CysSerGlyGluSerGluGlnLeuArgla...CysSerGlnAlaProCys 379  
 4949 GCGCGGATACACGCGCGCGCAAGAGGCTTCTGCTACCGAGCCTCGC 4900  
 379 sProGluGlnPro.....Aspp 386  
 4899 CGCGCCTACGCGCCCTGTGGCGCGCGAGAACGCGGCGTGACGGGGC 4850  
 386 roArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMetGlyGln 402  
 4849 CGCGGGAGACCGCGCGCGGGG.....CAG 4824  
 403 LeuTyGlnTrpGluPheProPheTrpGluVal.....GlnGlySerGlnAr 417  
 4823 ATCAGGAGGGGTGGCGCGCGCGCGCGCGCGCGCGCGGAGCGCGCG 4774  
 417 GcysGluLeuAsnCysArgProArgGlyPhe.ArgPheTyr..... 430  
 4773 GTCCCGCGCGGTACGGCGCGCGGGATCTCTCCGCGCTGGGCGGCTG 4724  
 431 .....ValAr 432  
 4723 TCCGCGCGCGCGCTCCCGCGGGGGGCGAGCACCCCGAGCGCGCGCG 4674  
 432 gHisThrGluLysValGlnAsp...GlyThrLeuCysGlnProGlyAlap 448  
 4573 CCAGCGCGACCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4625  
 448 roAspileCysValAlaGlyArgCysLeuSerProGlyCysAspGlyIle 464  
 4624 .....GTGGCGGTCGAGTGCCTGGCC...GCTGCGCGGGGATC 4589  
 465 LeuGlySerGlyArgArgProAspGlyCysGlyValCysGlyGlyAspAs 481  
 4588 CTGGAGCG.....GTGGCCGAGGCGCTTCGA 4563  
 481 pSerThrCysArgLeuValSerGlyAsnLeuThrAspArg..... 494  
 4562 CGCGGACCTGGCGGCGCTCCCGGGGCTGGCGGGCGCGCGCGCGCGCG 4513  
 495 .....GlyGlyProLeuGlyTyrGlnLysIleLeuTrpIlePro 507  
 4512 CCGCGCGCGGAGGACCGCGGGC.....CCC 4484  
 508 Alagly.....AlaLeuArgLeuGlnIle 515  
 4483 GCTTCCCGCGCGCGCGCGAGCGCGCGCGCGCGCTGCGCGCGTGCT 4434  
 515 eAlaGlnLeuArgProSerSerAsnTyrLeuAla..... 526  
 4433 GCGCGAGTCGCTGTCGCGGAGCCGCTGGTCTCATCGCGCTGCGCG 4384  
 527 .....LeuArgGlyProGlyArg..... 533  
 4383 GGGACTGCGCTGGCGCGCGAGCAGGCGCGGCGCGCGCGCGCGCGCG 4334  
 534 .....SerIleIle.AsnglyAsnTrpAla..... 541  
 4333 GCGGTAGCTGTGTCGCGGGGCGCTGGCGCGCGCGCGCGCGCGCGCG 4284  
 542 .....ValAspProPro..... 545  
 4283 GCGCTCGAGCTCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4234  
 546 .....GlySerTyrArgAlaGln 551  
 4233 ACCAGAGCTGGCGCCCTGCTGGGGGGCGCGCGCGCGCGCGCGCGCG 4184  
 551 yGlyThrValPheArgTyrAsnArgProProArgGluGluGlyGlyGly 568  
 4183 GCGAGCGCTGGCGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCG 4134

```
568 luSerLeuSerAlaGluGly..... 574
574 ..... 574
4133 CAAGCGCAAGAGTCCCGCGCCGCGCGCGCGCGCGAGCGCGCGCGCC 4084
574 ..... 574
4083 GACCCCGAGAGAGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCC 4034
575 .....ProThrThrGlnProValAspValTyrMetIlePheGlnGluG 589
4033 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4012
589 luAsnProGlyValPheTyrGlnTyrValIleSerSerProProProle 605
4011 .....CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3997
606 LeuGluAsnProThrProGluProValProGln..... 617
3996 CC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3947
618 .....LeuGlnProGluIleLeuArgValGluProProL 629
3946 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3897
629 euAlaProAlaProArgProAla.....ArgThrProGly 640
3896 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3847
641 ThrLeuGlnArgGlnValArgIleProGlnMetProAlaProProHisPr 657
3846 ACACGCGCGCGCG.....CCGCGCGCGCGCG.....CC 3821
657 oArgThrProLeuGlySerProAlaAlaTyrTrpLysArgValGlyHis 674
3820 CTGAGAGCCCTACT...GCTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 3774
674 erAlaCysSerAlaSerCysGlyLysGlyValTrpArgProIlePheLeu 690
3773 CCGCGCTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3751
691 CysIleSerArgGluSerGlyGluLeuAspGluArgSerCysAlaAl 707
3750 .....GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3731
707 aGlyAlaArgProAlaSerProGluProCysHisGlyThrProCysP 724
3730 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3681
724 roProTyrTrpGluAlaGlyGluTrpThrSerCysSerArgSerCysGly 740
3680 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3631
741 ProGlyThrGlnHisArgGlnLeuGlnCysArgGlnGluPheGlyGly 757
3630 ACGGCGCGCGCG.....GG 3617
757 yGlySerSerValPro.....GluArgCysGlyHisLeuP 770
3616 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3568
770 roArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysGlyHisTrp 786
3567 ..CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3520
787 GluValGlySerProTrpSerGlnCysSerValArgCysGlyArgGlyG 803
3519 .....TGCTGTACTCGCGCGCGCG..... 3502
803 naArgSerArgGlnValArgCysValGlyAsnAsnGlyAspGluValSerG 820
3501 ..CGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3453
820 luGlnGluCysAlaSerGlyProPro..... 828
```

```
3452 GTCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAACC 3403
829 .....GlnProProSerArgGluAlaCysAspMetG 839
3402 GGCTGTGCGCGCGCGCACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 3353
839 yProCysThrThrAlaTrpPheHis.....S 848
3352 CCCGACGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3303
848 erAspTrpSerSerLys.....ValSerProGlu 857
3302 GGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3253
856 ProProAlaIleSerCysIleLeuGlyAsnHisAlaGlnAspThrSerAl 874
3252 CCGCGGACG.....GGCGGCTCATCGTGGTGCACACCGGTG 3218
874 aPheProAla 877
3217 CGCGCGCGCG 3208

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq;US-08-458-568A-11

seq_documentation_block:
; Sequence 11, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,568A
; FILING DATE: 02-JUNE-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,146
; FILING DATE: 05-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn R.
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: DFCI-0029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Herpes simplex virus
; STRAIN: Herpes Simplex Virus Type 1
```

288 GlyTrpAlaSerProGlnValAlaGlyArgArgProAsnProPheProSe 304  
    |||     ::: |||     |||||     |||||     |||     |||  
2281 .....CTCCGCCCTCGCCCCCTCGCCCTCGCCCTCGCC...CCTCC 2318  
  
304 rValProArg..... 307  
    :::|||||  
2319 CCCCCTCGAATAAACACACTACTGCAAACTATCATCAGGTGTTCGCCG 2368  
  
308 .....GlyArgGlyGln 311  
    ::: |||  
2369 TTATTGGTCTTCGGGTCTCACAGCGCCCGCCCGTCCGCCCGCTT 2418  
  
312 GlnGlyGlnGlyProTrpGlyThrGlyGlyThrProHisGlyProArgLe 328  
    ::: |||     |||     |||||     |||||  
2419 ACAGACACCCTCCCTCGAACCGCGCGCTGCTTCGTCCCAAGCG 2468  
  
328 uGluProAspProGlnHisPro.....GlyAlaTrpLeuProLeuLeus 343  
    ::: |||     |||     |||||     |||||  
2469 CCTTCCAGTCCACAACCTCCCGCGCGGGGGCTGG..... 2505  
  
343 erAsnGlyProHisAlaSerSerLeuTrpSerLeuPheAlaProSerSer 359  
    ::: |||     |||     |||||     |||||  
2506 .....CCAAGCCGC 2514  
  
360 ProIleProArgCysSerGlyGluSerGluGlnLeuArgAlaCysSerGI 376  
    |||     |||     |||||     |||||  
2515 CTCCGCC.... 2523  
  
376 nAlaProCysProGluGlnProAspProArgAlaLeuGlnCysAlaA 393  
    |||     |||     |||     |||     |||     |||  
2524 ...CCAGCACTCCACGCGCCCGCGCGCGCCAGCA..... 2556  
  
393 laphAsnSerGlnLupHMetGlyGlnLeuGlyGlnTrpGluProPhe 409  
  
2556 ..... 2556  
  
410 ThrGluValGlnGlySerGlnArgCysGluLeuAsnCysArgProArgGI 426  
    |||     |||     |||     |||     |||     |||  
2557 .....CGGTGC..... 2562  
  
426 yPheArgPheTrValArgHisThrGluLysValGlnAspGlyThrLeuC 443  
  
2562 ..... 2562  
  
443 ysGlnProGlyAlaProAspIleCysValAlaGlyArgCysLeuSerPro 459  
    |||     |||     |||     |||     |||     |||  
2563 .....CGGTGC..... 2568  
  
460 GlyCysAspGlyIleLeuGlySerGlyArgArgProAspGlyCysGlyVa 476  
    |||     |||     |||     |||     |||     |||  
2569 .....GGCCGTGGCGCGAGCCGAC..... 2589  
  
476 lCysGlyGlyAspAspSerThrCysArgLeuValSerGlyAsnLeuThrA 493  
    ::: |||     :::  
2590 .....GAATCCGGCGCGCGC 2605  
  
493 sPArgGlyProLeuGlyTyrrGlnLysIleLeuTrpIleProAlaGly 509  
    ::: |||     |||     |||     |||     |||     |||  
2606 CGSGCGCAGGGCCCCCGGCGCTCGTCGCGCGCGACACCGAGCGGG 2655  
  
510 AlaLeuArgLeuGlnIleAlaGlnLeuArgProSerAsnTrpLeuAl 526  
    ::: |||     |||  
2656 GGCGCTCGT..... 2664  
  
526 aLeuArgGlyProGlyGlyVarSerIleAsnGlyAsnTrpAlaValA 543  
    |||     |||     |||     |||     |||     |||  
2665 .CGTCGGCTCCAGCAGGCGCG.....GGCGCA 2692  
  
543 spProGlySerTyrrArgAlaGlyGlyThrValPheArgTyrrAsnArg 559  
    |||     |||     |||     |||     |||     |||  
2693 AAAGTCCCTCCGCGCGCGCCACCGCGCGCGCGCGCG.....GCA 2736  
  
560 ProProArgGluGluGlyLvsGlvGluSerLeuSerAlaGluGlyProTh 576

|||||  
2737 CGGCTCGC.....GCCCCAGCCACG 2759  
576 rThrGlnProValAspValTyrMetIlePheGlnGluAsnProGly 593  
|||||  
2760 TACAGCGCGCGCA.....CGCGCGCGCCAGCGC 2788  
593 alPheTyrGlnTyrValIleSerSerProProIleLeuGluAsnPro 609  
2789 CCAGCGCGCGCGCGCGCGCGAGTGGCCCTCTCGCAGAGTCCG 2838  
610 ThrProGluProValProGlnLeuGlnProGluIleLeuArgValG 626  
2839 GCAGCGCGCGCGCGCGCGCGTGGTGGTCCCGAGG.....CC 2876  
626 uProProLeuAlaProAlaProArgProAlaArgThrProGlyThrLeuG 643  
2877 GCGCGCGCGCGCGCGCGCG.....CCGCGCAGCAGCGCGCGGTACTC 2920  
643 lnArgGlnValArgIleProGlnMetProAlaProHisProArgThr 659  
2921 GCGCGCGCGCATGGCAGCGCGGTGTCG..... 2949  
660 ProLeuGlySerProAlaAlaTyrTyrLysArgValGlyHisSerAlaC 676  
2950 .....GCGCGAGCGCGTG 2963  
676 s.....SerAlaSerCysGlyLysGlyValTyrArgProIlePheL 690  
2964 GCGCAGCGGTAGCGACGTGCGCGCGCG..... 2994  
690 euCysIleSerArgGluSerGlyGluGluLeuAspGluArgSerCysAla 706  
2995 .....ACAGCGCAGCGCGCGCGCGTGGGTACAGCG 3027  
707 AlaGlyAlaArgProAlaLaser..... 714  
3028 GCGGTGCGCGGCTCCACGCGCGAGACCCCGCGCGCGGCGGACAGCGG 3077  
715 .....ProGluProCysHisGlyThrProCysProProTyrTrpGluA 729  
3078 CCGAGCGCACACCGTGGCGCGCGCGCGTGGCGCGCGCG.....G 3118  
729 lagGlyGluTrpThrSerCysSerArgSerCysGly.....ProGly... 742  
3119 CCAGCGCAGCGCGCACTGCACGCGCGCGCAGCAGCTCGCAGCGCAGTAGG 3168  
743 .....ThrGlnHisArgGlnLe 748  
3169 CGTGTCCCGCAGACCGCGCGCGCGTGGCGCGCGCGAGTCGCGCGCGC 3218  
748 uGlnCys.....ArgGlnGluPheGlyGlyGly 758  
3219 ACGGTGTGACCAAGTACCGCGCGCGTGGCGCGCGCGCGCGCGCGCG 3268  
758 lySerSerValProGluArgCysGlyHisLeuProArgProAsnIle 774  
3269 CAGAACTCCAGCGCGCGCGCA.....AGCGCAGGTCCG 3303  
775 ThrClnSerCysGlnLeuArgLeuCysGly...HisTrpGluVal Gly 789  
3304 GCGTGGACAGCAGCAGCAGCGCTGTGGCGCGCGCGCGCGCGCGCGG 3353  
790 SerProTrpSerGlnCysSerValArgCysGlyArgGlyGlnArgSerAr 806  
3354 GCGCGGTCCCAATTCCCGCGCGCGCGCGGTGTCGCGCGCGCGCAGCGC 3403  
806 gGlnValArgCysValGlyAsnAsnGlyAspGluValSerGluGlnGluC 823  
3404 GTTGGCCAGGCGCGCGCAGCAGGAGGACACCGCGCGCGCGTGGCGGAC 3453  
823 ysAlaSerGlyProGln.....ProProSerArgGluAlaCysAsp 837  
.....

3454 ACTCGGGGGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3503  
838 MetGlyProCysThrThrAlaTrpPheHisSerAspTrpSerSer..... 852  
3504 ACGCGCGAGTACAGCACCCACCGCGCGCGCTCTCGGGGTGGGATCTG 3553  
853 .....LysValSerProGlu..... 857  
3554 GCGCATCCAGCGCGCGCATGCGCGCGAGCGCGCGCGCGCGCGCGCG 3603  
858 .....ProProAla 860  
3604 CAAAGAGCGCGCGCGCGCGCG 3623



Q9nsj8 homo sapien  
Q9w493 drosophila  
Q9ufz4 homo sapien  
P90884 caenorhabdi  
Q9vfe1 drosophila  
Q9wkv3 drosophila  
Q9xsv8 bos taurus  
Q75851 homo sapien  
Q9v9q7 drosophila  
Q9nrm2 homo sapien  
Q9cr7 mus musculu  
Q95875 homo sapien  
Q96qc6 homo sapien  
Q9lyp4 mus musculu  
Q9hsb6 homo sapien  
Q22631 caenorhabdi  
Q9cx59 mus musculu  
Q9uli7 homo sapien  
Q9h8x0 homo sapien  
Q9zlr1 mus musculu  
O41935 murid herpe  
Q9veg7 drosophila  
Q9uq1 homo sapien  
Q9tte5 bos taurus  
O61603 drosophila  
Q9p3j0 neurospora  
Q9bxj3 homo sapien  
Q90584 gallus gall  
Q99307 epstein-bar

17 466 9.5 550 4 Q9NSJ8  
18 460.5 9.4 1054 5 Q9W493  
19 414.5 8.5 898 4 Q9UFZ4  
20 403.5 8.2 1059 5 P90884  
21 391.5 8.0 1229 5 Q9VFE1  
22 357 7.3 566 5 Q9WKV3  
23 351.5 7.2 1637 6 Q9XSV8  
24 346.5 7.1 4123 4 Q75851  
25 325.5 6.6 858 5 Q9V9Q7  
26 318 6.5 432 4 Q9NRM2  
27 305.5 6.2 244 11 Q9CR7  
28 301 6.1 2157 4 Q95875  
29 301 6.1 2157 4 Q96QC6  
30 296 6.0 235 11 Q9LYP4  
31 292.5 6.0 237 4 Q9HSB6  
32 284 5.8 788 5 Q22631  
33 282.5 5.8 192 11 Q9CX59  
34 281.5 5.8 1023 4 Q9ULI7  
35 268.5 5.5 822 4 Q9H8X0  
36 268 5.5 2157 11 Q9ZLR1  
37 264.5 5.4 585 12 O41935  
38 264 5.4 2703 5 Q9VEG7  
39 262 5.4 364 4 Q9UCQ1  
40 260.5 5.3 685 6 Q9RTS5  
41 260 5.3 2715 5 O61603  
42 255 5.2 2592 3 Q9P3J0  
43 251 5.1 417 4 Q9BXJ3  
44 251 5.1 1146 13 Q90584  
45 250.5 5.1 924 12 Q99307

ALIGNMENTS

RESULT 1  
ID Q9EPX2 PRELIMINARY; PRT: 1280 AA.  
AC Q9EPX2:  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE PAPILIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1;  
RX MEDLINE=20530499; PubMed=11076767;  
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,  
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,  
RA Sieron A.L., Prockop D.J., Fessler J.H.;  
RT "Papilin in development; a pericellular protein with a homology to the  
RT ADAM's metalloproteinases";  
RL Development 127:5475-5485(2000).  
DR HSSP; P12111; 2KNT.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig-like.  
DR InterPro; IPR003006; Ig\_MMC.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR Pfam; PF00047; Ig; 3.  
DR Pfam; PF00014; Kunitz\_BPTI; 1.  
DR Pfam; PF00090; tspl; 5.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00409; IG; 3.  
DR SMART; SM00408; IGC2; 3.  
DR SMART; SM00410; IG-like; 2.  
DR SMART; SM00131; KU; 1.  
DR SMART; SM00209; TSPL; 5.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	738.5	15.1	1280	11 Q9EPX2	Q9epx2 mus musculu
2	731	14.9	3198	5 Q9D8G8	Q9d8g8 manduca sex
3	728.5	14.9	1235	4 Q95428	Q95428 homo sapien
4	714	14.6	2174	5 Q9GQR0	Q9gqr0 drosophila
5	711.5	14.5	3060	5 Q9VAV4	Q9vav4 drosophila
6	695.5	14.2	1572	5 O44938	O44938 haemochrom
7	662.5	13.5	454	4 Q96H81	Q96h81 homo sapien
8	661.5	13.5	538	4 Q9H8E4	Q9h8e4 homo sapien
9	640.5	13.1	2167	5 Q75840	Q75840 caenorhabdi
10	618	12.6	951	4 O60345	O60345 homo sapien
11	618	12.6	2165	5 Q19791	Q19791 caenorhabdi
12	532.5	10.9	1427	4 Q96L37	Q96l37 homo sapien
13	530	10.8	525	4 Q96RW4	Q96rw4 homo sapien
14	510.5	10.4	766	4 P82987	P82987 homo sapien
15	505	10.3	1014	5 Q95R33	Q95r33 drosophila
16	477.5	9.8	340	11 Q91Z56	Q91z56 mus musculu

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 21:04:46 ; Search time 67.92 seconds  
(without alignments)  
2233.754 Million cell updates/sec

Title: US-10-041-770-2  
Perfect score: 4895  
Sequence: 1 MENWTRPWLVLILLSLPQ.....PPAISCILGNHAQDTSAPPA 877

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

```
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.  
DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.  
DR PROSITE; PS00092; TSPI; 3.  
KW Immunoglobulin domain; Serine protease inhibitor.  
SQ SEQUENCE 1280 AA; 138824 MW; AE287705B561AF30 CRC64;  
  
Query Match 15.1%; Score 738.5; DB 11; Length 1280;  
Best Local Similarity 31.6%; Pred. NO. 3.3e-43;  
Matches 173; Conservative 50; Mismatches 226; Indels 99; Gaps 13;  
  
Qy 335 PCAWPLLSNGPHASSLSWLPAPSPPIR-----CSGSEQL 371  
Db 16 PSW---ARNVRQSDTWGTWGESPCSTCGGGSIFRERPCYSQRDRDGTSCVGPASRH 72  
  
Qy 372 RACSAQAPCPPEQDPALQCAAFNSQEFMGQLYOWEPTEVOGSGORCELNCRPRFRVY 431  
Db 73 RICHTEPCDGVDRPRAEQCAEFQDGTQGRYRWLPY--YAAPNKCENLIPKQGFY 130  
  
Qy 432 RHTEKVGDTLQCPADICVAGCLSPGCGILGSGRRPDGCGVGGDDSTCLVSNL 491  
Db 131 KHKDAVDGTCEPQQRDICVDGVRVVGCDHKLDSIKQDKLQCCGGGSCYPVTGTF 190  
  
Qy 492 TDRGGPLGYQKITLTPAGALRLQIAQLRPSSNYLALRGFGGRSIINGWAYDPFGSYRAG 551  
Db 191 DGNLDSRGYNQIIFIPAGATSIRIEAAASRNFLAVKSIRGEYILNGHWTIEAAQALPVA 250  
  
Qy 552 GVYFRYNRPREGGESLSAEGPTQPDVYVMFOENPGVYQYVVISPPPLENPTP 611  
Db 251 STVLYERGVEDLAPERLQARGPTSEPLVIELLSQESNPGVHYEYL----- 298  
  
Qy 612 EPPVQLOPEILRVPPPLAPAPRPARPTGTLQQRVRIQPMAPPHPRTPLGSPAAYKRV 671  
Db 299 -----PANDPGR-----GFSWSHG 312  
  
Qy 672 GHSASASCGKWWRPFLICISRESGELDERSCAAGARPASPEPCHGTCP--PYWEA 729  
Db 313 SWGDCSAECGGHGHQSLVFTTI--DNEAYPDHMCQHQR--PTHRSCNTQPCPKTKRWKV 369  
  
Qy 730 GEWTSRSCGPGTHRQLQCRQFEGGGSSVPPE--RGCHLP--RPNITQSCQLRLCGHWE 787  
Db 370 GPWTPCSVCGGGVQSRVYCLSSDGTGQEAETQACLAGKPTTQACNLQCAVWS 429  
  
Qy 788 VGSWQSVRCGRGQRSQVRCVNGNDEVSEQECASGPPPPSREACDMGCTT---- 843  
Db 430 V-EPWGECSVTGCTGIRKRSVTCRDEGSPVHAACLL--KDQPTLTQPCVQACPVFRGQ 487  
  
Qy 844 AMFHSWMS 851  
Db 488 AMHVGWS 495  
  
RESULT 2  
Q908G8 PRELIMINARY; PRT; 3198 AA.  
AC Q908G8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE LACUNIN PRECURSOR.  
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Spingilodea; Spingilidae; Sphinginae; Manduca.  
OX NCBI_Taxid=7130;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99457716; PubMed=10528409;  
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;  
RT "Expression of lacunin, a large multidomain extracellular matrix  
protein, accompanies morphogenesis of epithelial monolayers in Manduca  
sexta."  
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
```

```
DR HSSP; P12111; 2KNT.  
DR InterPro; IPR004094; Antistatin.  
DR InterPro; IPR003598; Ig_C2.  
DR InterPro; IPR003006; Ig_MHC.  
DR InterPro; IPR002223; Kunitz_BPTI.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR002221; WAP.  
DR Pfam; PF02822; Antistatin; 4.  
DR Pfam; PF00047; Ig; 2.  
DR Pfam; PF00014; Kunitz_BPTI; 9.  
DR Pfam; PF00095; wap; 1.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00408; IGC2; 2.  
DR SMART; SM00131; KU; 10.  
DR SMART; SM00209; TSPI; 7.  
DR SMART; SM00217; WAP; 1.  
DR PROSITE; PS00317; 4-DISULFIDE_CORE; 1.  
DR PROSITE; PS00280; BPTI_KUNITZ_1; 8.  
DR PROSITE; PS00279; BPTI_KUNITZ_2; 10.  
DR PROSITE; PS00092; TSPI; 1.  
KW Immunoglobulin domain; Serine protease inhibitor; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459CD09134 CRC64;
```

```
Query Match 14.9%; Score 731; DB 5; Length 3198;  
Best Local Similarity 28.5%; Pred. NO. 3e-42;  
Matches 179; Conservative 73; Mismatches 216; Indels 160; Gaps 17;  
  
Qy 305 VPRGRGQGGPWT-----GGTPHCPRLPEPDPQHPGAWLPLLSNGPHASSLSWLF 355  
Db 55 IPGEGE--TGWDGMDSTPCSRTCGGVASQKRI-----CUKFGPDQ----- 95  
  
Qy 356 APSPIPRCSGESQLRACQAPCPPEPQPPRALQCAAFNSQEFMGQLYOWEPTEVQGS 415  
Db 96 -----PQCTGGDTKYFSCETQDCPGSTDFRABQCSKYDDVEYRNKIKKPYT--RGP 147  
  
Qy 416 QRCELNCRPGRFVRYRHTKVDGTLQCPGADPICVAGRCCLSPGCGIILGSRPPGCG 475  
Db 148 NPELNCMPHGERYYRHQLKVIDGTRCNDDSFVGVNGTCQPVGCDMLGNSAREDKCR 207  
  
Qy 476 VCGGDDSTCLRVSNLDRGGLGYKILMTAGALRLQIAQLRPSSNYLALRGPPGRSI 535  
Db 208 KCRGNKCYTNTVLTQDLIKYNDMLLIPGATVIVMEVKPSNNYLALRAKNITYY 267  
  
Qy 536 INGNWAYDPFGSYRAGTVFRYNRPREGKESLSAEGPTQPDVYVMFOENPGVY 595  
Db 268 LNGDYHIDFPRSMIAGALWFYERSQQGFAAPDKRLCLGPTTEPLXLSLLQSNVNGIEY 327  
  
Qy 596 QYVISSPPPLENTPPEPPVQLQPEILRVPEPLAPAPRPARPTGTLQQRVRIQPMAPP 655  
Db 328 EYSL-----PTEAPRPTQN----- 342  
  
Qy 656 HPRTPLGSPAAYKRVGHSACSSACGKWWRPFLICISRESGELDERSCAAGARPASP 715  
Db 343 -----YTWVHEHTECSATCGGGGFQTRNVTCRQMDLELVDDTCLDEGLK--PVTN 391  
  
Qy 716 EPCGHTPCPPWEAGETSCRSOG--PQTHRQLQCRQFEGGGSSVPPER--CGHL--PR 771  
Db 392 QTCNNNVCPQWVTGEWCSCHRGSDGTOSREIQCKIIVNGPRSVISKRECFLLGPK 451  
  
Qy 772 PNITQSCQLRL--CGHWVEYGVSPWQSVRCGRQSRQVRVGVNNGDEV----- 818  
Db 452 PEKFRKCNVNTCTPTWTG--PWKPCDLYGBGKQTRQVVCVQKNRRRIDVLDSDSECTDER 510  
  
Qy 819 --SQEC-----ASGPPPPS 832  
Db 511 PESEOKCMIHSEERTDWVASEWSGDCNCFSTMRTRIAKTYDRKLVDVSFCAHPLPVL 570  
  
Qy 833 REACD---MGPCTTAWFHSWSSKVSPE 857  
Db 571 QEPDAATLPPCDVQWYATQW--SKCSSE 597
```

```
RESULT 3
Q95428 PRELIMINARY; PRT; 1235 AA.
AC O95428;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 133.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Madan A., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A.,
RA Madan A., Dickhoff R., Shaffer T., James R., Lasky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109907; AAC97963.1; -
DR HSSP; P12111; 2KNT.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; Ig_3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00092; TSP1; 4.
KW Hypothetical protein; Immunoglobulin domain;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 14.9%; Score 728.5; DB 4; Length 1235;
Best Local Similarity 30.7%; Pred. No. 1.6e-42;
Matches 182; Conservative 51; Mismatches 229; Indels 131; Gaps 16;

QY 299 PDPFVPRGRGQGGQ-GPWGT-----GGTPHGRPLEDPQHPGAWLPLLSNGPHA 348
DB 13 PAPGSSAPKVRGRSDTWGFWQSPSCRTCCGGVSPFR-----PCYSQRDGG 60
QY 349 SSLWSLFASSPTPRCSGSEQLRACSQAPCPPEQDPALQCAAFNSOEFMGOLYOWEP 408
DB 61 GS-----SCVGPASHRSRTSCPDGARDFAEQCAEFDGAEPGGRYRWLP 108
QY 409 FTEVQGSQRCCLNCRPGRFRFYVHTEKVDGTLCPQGPAPDICVAGRCLSPGCDGILGSG 468
DB 109 Y--YAPNKECLNCLIPKGENFYKHEAVVDGTPCEPGKRDVCDGSCRVWVGCDHELDSS 166
QY 469 RRPDGGVCGGDSTRLVSGNLTDRGGPLGYOKILWIPAGALRLQIAQLRPSSNVALR 528
DB 167 KQEDKCLRCGGDGTCTCPYVAGTDFDANDLSRGYNQIILVPMGATSIILIDEAASRNFLAYK 226
QY 529 GPGRSIIINGNAVDPPGVSFRAGTFRYRNPPREGKGSLSAEGTTPQVDVYMFQEZ 588
DB 227 NVRGYYLNHGWTHIARALPAASTILHYERGAEGDLAPRLHARGTSEPLVIELISQE 286
QY 589 ENPGVFYQYVSSPPPILENPTPEPPVQLQPEILRVPPPLAPAPARTPTGTLQKVRI 648
DB 287 PNPGVHVEYHL-----PLRRP----- 302
QY 649 PQMPAPHPRTPLGSPAAYWKRVRGHSACSASCGK-VWRPIFLICISRESGEELDERSCAA 707
DB 303 -----SPGFSHGSWSDCSAECGGGHOSRLVFTCTIDHEA---YPDHWCQR 345
```

```
QY 708 GARPPASPEPCHGTGTPCP-----YWEAGETSCSRSCGPGT 743
DB 346 QPR-PADRRSCNLHPCPEFKRTSYLHRPGAWRLAGAVCGNSWKAGPWAPCASCGGGS 404
QY 744 QHROLQCRQEFEGG-GSSVPPERCCHLP-RPNITQSCQLALCGHWEVGPSPWSCSVRCGR 801
DB 405 QSRSVYCISSDAGAGIOEAVEAEACAGLPKPPAIAQCNLQRCAAWS-PEFWGECVSVCV 463
QY 802 GQSRQVRCVGNNGDEVSEQECASGPPQPPSREACDMGQC-----TTAWFHSDW 850
DB 464 GVKRSVTCRGERSLLHTAAC-SLEDRPPLTEPCVHEDCPLLSDQAWHVGTV 515

RESULT 4
Q9GQRO PRELIMINARY; PRT; 2174 AA.
AC Q9GQRO;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EXTRACELLULAR MATRIX PROTEIN PAPILIN PRECURSOR.
GN PPN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
RA Sieron A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RT ADAMTS metalloproteinases."
RL Development 127:5475-5485(2000).
DR EMBL; AF205357; AAG37995.1; -
DR HSSP; P12111; 2KNT.
DR Flybase; FBgn0003137; ppn.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig-like.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00014; Kunitz_BPTI; 3.
DR PRINTS; PR00003; 4DISULPHCORE.
DR PRINTS; PR00759; BASICPTASE.
DR SMART; SM00409; IG; 3.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00410; IG.Like; 1.
DR SMART; SM00131; KU; 3.
DR SMART; SM00209; TSP1; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 3.
DR PROSITE; PS00092; TSP1; 3.
KW Matrix protein; Serine protease inhibitor; Signal.
FT SIGNAL 1 26
SQ SEQUENCE 2174 AA; 231936 MW; 03BF707952623120 CRC64;

Query Match 14.6%; Score 714; DB 5; Length 2174;
Best Local Similarity 29.7%; Pred. No. 3e-41;
Matches 174; Conservative 85; Mismatches 209; Indels 118; Gaps 21;

QY 302 FFSVPRGRGQGGGW--GTGCTPHGRPLEDPQHPGAWLPLLSNGPHASSLSLFAFS- 358
DB 28 FPGFL-RQKQYGANMILPESVTPGEGNDPD-----EWTP-----WS--SPSD 68
QY 359 -----SPIPR-----CSGESQLRACSOAPCPPEQDPALQCAAFNSQ 397
DB 69 CSRTCGGVSYQTRCCLRRDRGEAVCGSGSRRTSCNTQDCPEESDFRAQCQSRFDQ 128
```

QY 398 EFMGLYQWFFTEVOGSCRELNCRRGRFFVYHTEKVDGTLQCPGAPDVCVAGRL 457  
 DB 129 QFDGVYFVWVPTNAPN-CELNCMPAGERFYRQREKVVGDGFRCKNDKLDVCGNECM 186  
 QY 458 SPGCDGILSGRRPDDGCGVCGDSDTCLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQ 517  
 DB 167 PVGCDMLGSDAKEDKCRKCGDSDTCTIRNTITKNLAPGYNDLLLPAGATNIRIEE 246  
 QY 518 LPSSNYLALRGPGRRSTINGNWDVPGSVRAGTVPFRYRNPREECKGSLAEGETT 577  
 DB 247 TVPSSNYLACRHHGYYLNGDWRIDFRPMFFANSMWYORKPMGFAAPDLTCSGPI 306  
 QY 578 QPVDVYMFQENPGVYQYVYISSPPPILENPTBPVPPVLPQLPELRLVPPPLAPAPR 637  
 DB 307 ESLFVLMVQEKNSLDYEYSI-PE-----SLSHSQDTH 340  
 QY 638 TPGTLQVRIPQMPAPPHPRTPLGSPAAYKRVGHSCASCSGCGVWRPFLICISRESG 697  
 DB 341 T-----WTHQFNACSCSGGSONRKYTCNNRITL 371  
 QY 698 EELDRSCAAGARPASPEPCGHTPCPPYWEAGWTSCRSRG-PGTQRLQRCRQFEG 756  
 DB 372 AEVNSLQDQSK-PVEQACGTETECAPHWVEGWSKCGSGDGFQNRITCERISS 430  
 QY 757 GSSVPPPE-----RCGHLPRPNITOSC--QLRLCGHVEGSPWSOCVRCGRGQSRQVR 809  
 DB 431 GEHTVEEDAVCLKEVGN--KPAKQECNRDVKNCPKYHLG-PWTPCDKLCGDGKQTRKVT 487  
 QY 810 C-VGNNGDF--VSEOCASGPPQPSRACDMGPCT-TAFHSDWS 851  
 DB 488 CFIEENGHRVJPEEDCVE--EKPTKESCLTTCGVDWIIQSWS 531

RESULT 5  
 ID Q9VAV4 PRELIMINARY; PRT: 3060 AA.  
 AC Q9VAV4: Q9VAV3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE CG1540 PROTEIN.  
 GN PPN OR CG1540 OR CG18436.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Phylloidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktarolu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Warkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,  
 RA Shue B.C., Sidenkianis I., Stapleton M., Strong R., Sun E.,  
 RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Svirska R., Tector C., Turner R., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC EMBL; AE003765; AAF56794.2; -;  
 CC EMBL; AE003765; AAF56795.2; -;  
 CC HSP; P12111; 2KNT.  
 CC FlyBase; FBgn0003137; Ppn.  
 CC InterPro; IPR000561; EGF-like.  
 CC InterPro; IPR003598; Ig-C2.  
 CC InterPro; IPR003006; Ig\_MHC.  
 CC InterPro; IPR002223; Kunitz\_BPTI.  
 CC InterPro; IPR000884; TSPI.  
 CC InterPro; IPR002221; WAP.  
 CC Pfam; PF00047; Ig; 3.  
 CC Pfam; PF00014; Kunitz\_BPTI; 12.  
 CC Pfam; PF00090; tsp\_1; 5.  
 CC Pfam; PF00095; wap; 1.  
 CC PRINTS; PR00759; BASICPTASE.  
 CC SMART; SM00408; IGC2; 3.  
 CC SMART; SM00131; KU; 12.  
 CC SMART; SM00209; TSPI; 7.  
 CC SMART; SM00217; WAP; 1.  
 CC PROSITE; PS00317; 4-DISULFIDE CORE; 1.  
 CC PROSITE; PS00280; BPTI\_KUNITZ\_1; 11.  
 CC PROSITE; PS00279; BPTI\_KUNITZ\_2; 12.  
 CC PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 CC PROSITE; PS50092; TSPI; 3.  
 CC Alternative splicing: Immunoglobulin domain;  
 KW Serine protease inhibitor.  
 FT VARSPLIC 2803 2803 L -> SVVPV (IN SHORT ISOFORM).  
 FT VARSPLIC 2844 2854 FNEKTMDSGI -> VASPLHPNAV (IN SHORT  
 FT VARSPLIC 2855 3060 MISSING (IN SHORT ISOFORM).  
 FT VARSPLIC 3060 AA; ACA31D3E558C7C0 CRC64;  
 SQ SEQUENCE

Query Match 14.5%; Score 711.5; DB 5; Length 3060;  
 Best Local Similarity 31.4%; Pred. No. 6.4e-41;  
 Matches 157; Conservative 78; Mismatches 190; Indels 75; Gaps 14;

QY 364 CSGSEQLRACQAPCPPEQDPDRALQCAANSQFPMQLQWEPFFTEVOGSCRELNCR 423  
 DB 55 CSGGRRYFSCNTQCPPEESDFRAQCSRFDRQDFGVFVWVPTNAPN-CELNCM 112  
 QY 424 PRGFRFYRHTKVDGTLQCPGAPDVCVAGRCISPGCDGILSGRRPDDGCGVCGDSDT 483  
 DB 113 PKGERFYRQREKVVGDGFRCKNDKLDVCGNECMVPGCDMLGSDAKEDKCRKCGDST 172  
 QY 484 CRLVSGNLTDRGGPLGYQKILWIPAGALRLQIAQLRPPSSNYLALRGPGRRSTINGNWD 543  
 DB 173 CKTIRNTITKNLAPGYNDLLLPAGATNIRIEFVPPSSNYLACRHHGYYLNGDWRID 232  
 QY 544 PPGSVRAGTVPFRYRNPREECKGSLAEGETTQPVDVYMFQENPGVYQYVYISSPP 603  
 DB 233 FPRPMFFANSMWYORKPMGFAAPDLTCSGPISSSLFVLMVQEKNSLDYEYSI---- 288  
 QY 604 PILENPTPEPPVLPQLPELRLVPPPLAPAPRPTPGTLQRCVRIPQMPAPPHPRTPLGS 663

Db 289 -----PE-----SLSHSQDTH----- 301  
QY 664 PAAYKRYGHSASCSGKGYVRPIFLICISRESGEELDERSCAAGARPPASPECHGTPC 723  
Db 302 -----THHQFNACSGSGGSGSRKVTNNRITLAENPNSLCDOKSK-PVEEQACGTEPC 356  
QY 724 PYPWAGEWTSRSCSG-PGTQHRLOQCRQBFGGGSSVPPE-----RCGHLPRPNITQS 777  
Db 357 APHWVEGWSKSGSGSGDFQNRISITCERISSGHEITVEDAVCLKEVGN--KPAIKQE 414  
QY 778 C--QLRLCGHVEGSPWQCSVRCGRGORSQVRC-VGNNGDE--VSRQECASGPPPPPS 832  
Db 415 CNRDVKNCPKHYLG-PWTPDKLKGDKQTRKVTCTFIENGHGRVLPEDCDE--EKPET 471  
QY 833 REACDMGPCT-TAWPHSDWS 851  
Db 472 EKSCLLTCEGVDMWLSQMS 491  
RESULT 6  
ID O44938 PRELIMINARY; PRT; 1572 AA.  
AC O44938;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DI 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE THROMBOSPONDIN.  
GN THRI.  
OS Haemochus contortus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;  
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.  
OX NCBI\_TaxID=6289;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOREDUN;  
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,  
RA Knox D.P.;  
RT "Cloning and characterization of Thrombospondin, a novel multidomain  
RT glycoprotein associated with the gut of Haemonchus contortus";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043121; AAB99830.2; -;  
DR HSSP; P05067; ICA0.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR000884; TSPI.  
DR Pfam; PF00014; Kunitz\_BPTI; 12.  
DR Pfam; PF00090; tsp\_1; 6.  
DR PRINTS; PR00759; BASICPTASE.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00131; KU; 6.  
DR SMART; SM00209; TSPI; 7.  
DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 6.  
DR PROSITE; PS00092; TSPI; 2.  
KW Serine protease inhibitor.  
SQ SEQUENCE 1572 AA; 171871 MW; 2260830DC2F903EC CRC64;

Query Match 14.2%; Score 695.5; DB 5; Length 1572;  
Best Local Similarity 28.3%; Pred No. 4e-40;  
Matches 184; Conservative 70; Mismatches 230; Indels 167; Gaps 18;  
QY 220 SVHTPSQAEPLSPETAQTEVAPTRPAPLRHHPRQAASGTPEPSTHSLGEGGFRAFP 279  
Db 24 NVEPTLPHLPLSP-----PETNPAGAR----- 46  
QY 280 QPAPSSOGHASPQVAGRDPDPSPVPRGR-GQGGGPGWGT-----GGTPHGPR 327  
Db 47 -AKRAQVQTV-----DGSVSLDKTGSGSGPGWPVPEQCSRTCGGVQTEKR 96  
QY 328 LEPPDQHPGAWLPLLSNGPHASSLWSLAFSPSPIPRCSEGEQLRACSAQAPCPPEQDPR 387  
Db 97 -----QCPG-----DCTGSPVRYVSCNLEPC-PEGTDFR 124

QY 388 ALQCAAFNSQEFMGQLYQWBEFTTEVQSGRCELNCRRPRFRFRFVVRHTEKVDQTLCPGA 447  
Db 125 AEQCAAHNDPIDQGYHKWIPY---KGKNCCELLCKPENGFYKWDQTVVDGTCDSKG 181  
QY 448 PDICVAGRCLSPGCDGTLGSSRRPDGCGGDDSTCRVLSGNLTDGSGPLGQKILWIP 507  
Db 182 DDICVDGCVLPLGCDGKLSALKLDKCGVCDGSGSQCKTIEGSDERNLSFGYHDVWIP 241  
QY 508 AGALRLQIAQLRFPSSNYLALRPGGKRSIINGNAVDPGSGYRAGTVFRYRNPREEKG 567  
Db 242 AGATAIRIEARPPSSNNLAKMNSDDYFLANGSMIOVEKDVNLGVHFY-----DDGKP 296  
QY 568 ELSIAEGPTTQPDVDMYMFQENPGVYQVIVSSPPILNPTPEPPVQLQPELLRVEP 627  
Db 297 ERITAKGLEEEVVSVLIRKGRNDSIKYEFS----- 329  
QY 628 PLAPAPRPARTPGTLQQRVPIQOMPAPPHPTPLGSPAAYKRWVGH-SACASCSGKGVWR 686  
Db 330 -----VPIDVEDVMYKPGEWSACSVCCKGQQT 358  
QY 687 PIFLCISRESGEELDERSC-AAGARPPASPECHGTPCPYWEAGETSCSRSCG-PGTQ 744  
Db 359 RTPYCIETQARVDDQLDDANSTKPEFEKPCETVDCQAEWFKGEWPCSCGCGGGEQ 418  
QY 745 HROLQCRQEFGGGSSVPPERCGLHPRPNITQSCQLRLCGHVEGSPWQCSVRCGRGQR 804  
Db 419 YRVYCHQVFANGRRITVDDGNCCTAERPVRQVRCNRFACPEWQAG-PWSACSEKCGDAFQ 477  
QY 805 SRQVRC-----VGNNGDEVSEQECASGPPQPPREACDMGPCT-TAWPHSDW 850  
Db 478 YRSVTCSEKGEERGKLLPAPACSS-ENTLESQRSCNLGPTGLKFFTEW 527  
RESULT 7  
Q96H81 PRELIMINARY; PRT; 454 AA.  
ID Q96H81  
AC Q96H81;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DI 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 51.1 KDA PROTEIN.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC008840; AAH08840.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 454 AA; 51054 MW; A674483B13A18F93 CRC64;

Query Match 13.5%; Score 562.5; DB 4; Length 454;  
Best Local Similarity 39.4%; Pred. No. 2.1e-38;  
Matches 132; Conservative 45; Mismatches 121; Indels 37; Gaps 6;  
QY 554 VRYNRPR-BEGKESLSABGPTTQPDVDMYMFQENPGVYQVIVSSPPILNPTPE 612  
Db 1 MetyKRPNEISSTAGESFLAESPTNEILDVMIHQQPNPGVHYEYVINGTNAI----SPQ 56  
QY 613 PPVQLQP-----EILRVEPPLAPAPRPARTPGTLQQR 645  
Db 57 VP-PPRRPEPFNGOMVTGKRSQEGEKGNEEKEDURGEAPENFTSESQTPFVRHPD 115  
QY 646 VRIPOMP---APPHPRTPLGSPAAYKRWVGHSAASCSGKGVWRPIFLICISRESGEELDE 702  
Db 116 RSPHPNPLVPAPQPPRRSDHWNKQLGTETCSTCTGCKGQYPIPRCVHRSHTERAPE 175  
QY 703 RSCAAGARPPASPECHGTPCPYWEAGETSCSRSCGPGTQHRLOQCRQEFGGGSSVP 762





DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
DR PROSITE: PS50092; TSP1; 6.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN1.  
SQ SEQUENCE 2165 AA; 244337 MW; FCC3DA8AA9C4888 CRC64;

Query Match 12.6%; Score 618; DB 5; Length 2165;  
Best Local Similarity 26.5%; Pred. No. 1.4e-34;  
Matches 190; Conservative 90; Mismatches 254; Indels 184; Gaps 30;

QY 316 PWGTGTP-----HGP-RLEPD-PQHPGAMLPILNSGPHASSLW-----SLFA 356  
DB 571 PW-ADGTPCDESRSNMFCHGACVRLAPESLTKIDQWGDWRSWGECSRTCGGGVOKGLRD 629  
QY 357 PSSPIPR-----CSGESQLRACQAPCPPEQDPQPRALQCAAFNSQEP-----MGQLYQW 406  
DB 630 CDSKPRNGKVCYQORERYSCNTQECPWDTPYREVQCSEFNKIDIGIOGVASTNTHW 689  
QY 407 EP-FTEVGSQRCLENCPRGRFYVRTEKVDGTLCOPADICVAGRCLSPGCGDIL 465  
DB 690 VKYANVAPNERCKLYCRLSGAAYFYLLRDVKGDTGCDRNGDDICVAGACMPAGCDHQL 749  
QY 466 GGRPDGCGVCGGDDSCRLVSGNLTRDGGPLGYQKILWIPAGALRLQIAQ-----LRP 520  
DB 750 HSLTRDKCGVCGDDSKVVKTFNQ-GTFGYNEVWKIPAGSANIDIQKGYNNMKE 808  
QY 521 SNNYALRGPGSRSTINGNAVDPGYSYRA-GGTVFRYNRPRREGKSGESLSAEGTTPQ 579  
DB 809 DDNYLSLAANGFEFLNGHFVLSARQOIAFDQTVLEYS---GSDAIIERINGTPIRSD 865  
QY 580 VDVMYI-FQENPGVGYOYVTSISPPILENTPTPEVPVQLQPEILRVEPPLAPAPRPT 638  
DB 866 IYVHVLVSGHPDYSIYMTAAVP---NAVIRISSALY--LWRVTDWTECDRACRG 919  
QY 639 P-----GTQLR 644  
DB 920 QOSQKWLCLDMSTRQSHDRNCQNVLPKPKQATRMGNIDICSTRWITDVSSCAKCGSQK 979  
QY 645 QVRIP-----QMPAPPH-----PRTLGSFAAY-----WKEVGHSAASAG-KGV 684  
DB 980 RQVSCVMKMGDRQTPASEHLCDNRNKSFDIASCVIDCSGRKNYGEWTSCTSETCGSNGK 1039  
QY 685 WRPIFLCSRESGEELDRSCAAGRAPSPAPCPCHGTGTPPYWEAGETWTSRSCGRGTQ 744  
DB 1040 MHRKSYCYVD-DSNRVVDSELCGR-EQKEATERECNRIPC-PRWYVGHWSERSCSDGVK 1096  
QY 745 HRQLQCRQ-----EFG-----GGG-----758  
DB 1097 MRHAQDLAADRTHTSRCGPAQTOEHCHNEHACTWQFVWSDCSAKCGDGVQYRDANCT 1156  
QY 759 -----SSVPPERCGLPRNITQSCOLRLCGHWEVGSQSVRCGRQGRQVRCVGN 814  
DB 1157 DRHRSVLPEHRLKMEK-IITKPCRESCKPKLGE-WSQSVSCEDGWSRRVSCVSGN 1214  
QY 815 GDEVSEQACGPPPPSRACDMPCTTANFHSWSSKVSPEPPAISCILGNAQDT 872  
DB 1215 GTEVDSMLCGTASDRPASHOTCNLGTG-CFFWRNTDWSA-----CSVSCGIGHRETT 1265

RESULT 12

Q56L37 PRELIMINARY; PRT: 1427 AA.  
ID Q96L37  
AC Q96L37;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VON WILLEBRAND FACTOR-CLEAVING PROTEASE PRECURSOR.  
GN ADAMTS13.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX PubMed=11557746;  
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,  
RA Fujikawa K.;  
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a  
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";  
RL J. Biol. Chem. 276:41059-41063(2001).  
DR EMBL: AV053376; AAL17652.1; -;  
KW Signal; Protease.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.  
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCA14442 CRC64;

Query Match 10.9%; Score 532.5; DB 4; Length 1427;  
Best Local Similarity 23.5%; Pred. No. 8e-29;  
Matches 272; Conservative 86; Mismatches 430; Indels 371; Gaps 50;

QY 19 PQLCLDQEVLSGHSLOTPTEEGQGPVGWGFVQWASCSQPCGCVGVRRTQCLPTVOL 78  
DB 11 PPLCVAGILACGFL-----CGWGPSHFQOCLQ-----ALEPQAVSSYL 50  
QY 79 HPSLPLPPRPRHPEALLPRGQPRPTSPETLPLVXTOSRGRGGLR-----GPASH 131  
DB 51 SPQAPLKGRRP-----SPGFQQRQRRRAAGGILHLELLVAVGPDVF 93  
QY 132 LGREETQEIARAAR-----SRLRDPKPGMGYGRYPFALPLHNRHRHPRSPRSELSLI 186  
DB 94 QAHQEDTERVLTMLNIGALLRDPISLGAQFRVHLVKWVI-LTEPEGAPNITANLTSSL 152  
QY 187 SSRGEEAIPPTPRAPFSAN-----GSPOTELPPTLSVHTSPQAEPLSPE-----234  
DB 153 SVCGWSOTIPEDTDPGHADLVLYTRFDLELPGNRQVRGVYTLQGGACSPWSCLITE 212  
QY 235 -----TAQTEV-----APTRPAPLRHHPRAQASGTEPP-----SPTH---267  
DB 213 DTGFDLGVTTAHEIGHSGFLEHGDAGSGCGPSGH---VMSDGAAPRAGLAWSPCSRQ 269  
QY 268 --SLGEGGFTRASPQPRPSSQGWASPOVAGRDPDPFSPVPRGQGGQ---CPWGTGTT 322  
DB 270 LLSLLSAGRARCVMDPRP-----QFGSAGHPDPAQGLYYSANEQCRVAFGFKAVACT 323  
QY 323 PHGPRLE-----DPQHPGA-----WLPLL-----SNG-----PHAS-349  
DB 324 FAREHLDQALSCHTDPLDQSSCSLLVPLLDGTGCGVEKWCCKGRCSLVELTPIAAV 383  
QY 350 -SLWSLFAPSSPIPR-----CSGESEQLRACSQAPCPPEQP 384  
DB 384 HGRWSSWGRSPSRSCGGGVVTRRRQNNPRPAFGGRACVADLQAEMLTQACEKTQL 443  
QY 385 DPALQCAAFNSQEFM-----GOLYQW-EPTFVQGSQRCLENCPRGRFRFYVRHTEKVQ 438  
DB 444 EFWSQCARTDGQLRSSPGGASFYHNGAAPHISQGDALCRHMCRAIGESFIMKRGDSFL 503  
QY 439 DGTLCQPGAP-----DICVAGRCLSPGCGILGSGRRPDGCGVCGDDSTCLVSGNLT 492  
DB 504 DGTFCMPSGPRDGTLSLCVSGSCRTFGCDGRMDSQVMDRCQVCGGDNSTSPRGSEFT 563  
QY 493 DRGGLGYQKILMIPAGALRLQIAQLRPPSSNYLALRGPGRSIINGNAVDPGYSYRA---550  
DB 564 -AGRAREYVTLVTPNLTSTVYIANRPLETHLAVR-IGRVYVAGKMSIPNTTPSLL 621  
QY 551 --GGTVFRY-----NRPPREGKSGESLSAEGTTPQVD--VYMIFQEE-----NPGVFYQY 597  
DB 622 EDGRVEYRVALTEDRLPRL-----EERINGPQEDADIQVYRRYGEYGNLTPDITFY 677  
QY 598 VISSP-----PP-----604  
DB 678 FQKPRQAWAAVAVRGPCSVSCGAGLRWNYSCLDQARKELVETVOCQSQQPPANPEAC 737  
QY 605 ILENPTP-----EPPV--PQLQPEILRVEPFLA-----630  
[1]



Db 738 VLE-PCPPYAVGVDFGSCASCGGGLRERPVKVCVEAQGSLKLTLPAPCRAGAQAVAL 796  
QY 631 ----PAPPAR-----TPGLQORVRIP 649  
Db 797 ETCNFQPCPARVESEPSSTAGGAGLALENETCVPAGDGLAPVTEGPGSVDEKLPAP 856  
QY 650 Q----WPAPP-----HPPTPLGS-----PAAVWKRGVHGSACSACGKGVWRP 587  
Db 857 EPCVGMSCPPGNGHLDATSAKAPSPWGSINTGAQAARVWTPVAGS-CSVSCGRGLMEL 915  
QY 688 IFLCISRESGEBLERSCAAGARPPASPECHGTGTPPYWEAGETSCSRSCGPGTQHRQ 747  
Db 916 RELCHDSALRPVPELGLASKPSRREVCOAVCPARWQY-KLAACSVSGRGVVERI 974  
QY 748 LQCRFEG--GGGSSVPPEPCGHLRPNITQSCURLC-GHWEVGSPPWSQSVRCGRQOR 804  
Db 975 LYCARAHGDDDEEILLDTQCQGLRPERPEQAECSLEPCPRWKVMS-LGPCSASGLGTA 1033  
QY 805 SQVRCVG--NNGD---EVSROEACASGPPPPSREACDMGPTTAWPHSDSWSSKVSPEPAI 861  
Db 1034 RSVACVQLDQDQVEVDEACAA-LVRPEASVPCLADCIYRWHVGTME-----CSV 1086  
QY 862 SC-----ILGNHAQ 870  
Db 1087 SCGDGIQRRRDCGLGPAQ 1105  
  
RESULT 13  
Q96RW4 ID Q96RW4 PRELIMINARY; PRT; 525 AA.  
AC Q96RW4;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ADAM-TS RELATED PROTEIN 1.  
GN ADAMTSR1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hirohata S., Apte S.S.;  
RT "A novel member of ADAM-TS related gene, ADAM-TSR1 (A Disintegrin-like  
RT And Metalloproteinase domain with Thrombospondin type I modules  
RT Related gene-1).";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF176313; AAK84170.1;  
SQ SEQUENCE 525 AA; 58351 MW; 3F180C3CAA7BA68 CRC64;  
  
Query Match 10.8%; Score 530; DB 4; Length 525;  
Best Local Similarity 25.9%; Pred. NO. 4e-29;  
Matches 149; Conservative 61; Mismatches 213; Indels 152; Gaps 18;  
  
QY 351 LWSLFAPSSPIPR-----CSGSEQLRACSQAQPCPEQDPDRALQCA 392  
Db 35 LWDAGFWSECSCTCGGGAANSURCLSKSCGGRNIRVTCNSVNDPCPEAGDFRAQCS 94  
QY 393 AFNSQFMGOLYQWEPFTEVQGSQRCNCRPRGFRFYVHTEKVDGTLCPQGPAPDVCV 452  
Db 95 AHNDVKHGGFYEWLPVSN-DPDNPSCLCAQAKGTLVLVAPKVLDTGRCVTSLSDMCI 153  
QY 453 AGRCISPGCDGILGSRRRPDGGVCGGDSTCLVSGNLTDR-GGPLGQKILNTPAGAL 511  
Db 154 SGLCQIVGCDHQLGTVKEDNCVCNGDGTCLRLVRGQYKSLSATKSDTVAIPIGSR 213  
QY 512 RLQIAQLRPSNYL---ALRGPGRSIIINGNAVDPGGSYRAGGVFRYRNRPRREGKE 568  
Db 214 HIRLVKGPDLVLETKTLQIKGENSLST-----GTFVLDNSVDVQKFPDK-----E 263  
QY 569 SLSAEGTTPQVDVYVIFQENPG-----VFYQYVISSPPILENTPEPPVPLQ 620

Db 264 ILRWAGPLTAD----FVIVIRNNGSGADSTVQVIFYQ-----PIIHR----- 300  
QY 621 EILRVPEPLAPAPRPARTPGTLQORVIPOMPAPPHPTPLGSPAAAYWKRGVHGSACSASC 680  
Db 301 -----WRETDFFPCSATC 313  
QY 581 KGWVRPIFLICISRESGEBLERSC---AAGARPPASPECHGTGTPC----- 724  
Db 314 GGGYQLTSAECYDLRSNRVADQYHYYPENIKPKLQECNLDPCPASDGYKQIMPYDL 373  
QY 725 ---PYWEAGETSCSRSCGPGTQHRQLC-ROBFGGGSSVPPPERGCHLPRPNITQSCQ 779  
Db 374 YHPLPRWEATPWATCSSCGGIGSRACVSEVEDIQGHVTSVEWKCMYTPKMPIAQPCN 433  
QY 780 LRLCGHVEVSPWSQSVRCGRQORSVRCVGNNGDEVSEQECASGPPPPPREACDM- 838  
Db 434 IFDCPKW-LAQEWSPCVTGCGGLRYRVLICIDIRGMHTG---CSPKTPHKEECIVP 489  
QY 839 GPC-----TTAWFHS-----DMSKSVSPRP 858  
Db 490 TPCYKPKLEKLPVEAKLPWFKQAQLEEGAAVSEEP 524  
  
RESULT 14  
P82987 ID P82987 PRELIMINARY; PRT; 766 AA.  
AC P82987;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE ADAM-TSL3 PRECURSOR (FRAGMENT).  
GN ADAMTSL3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hirohata S., Anand-Apte B., Seldin M., Apte S.;  
RT "Punctin, a member of a new family with similarities to ADAM-TS  
RT proteases, is a component of extracellular matrix of skeletal  
RT muscle";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.  
DR EMBL; AF237652; AAK15041.1;  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp.1; 6.  
DR SMART; SM00209; TSP1; 7.  
DR PROSITE; PS50092; TSP1; 1.  
KW Repeat; Signal;  
FT SIGNAL 1 26  
FT CHAIN 27 766  
FT DOMAIN 79 123  
FT TSP TYPE-1 1.  
FT DOMAIN 422 474  
FT TSP TYPE-1 2.  
FT DOMAIN 482 528  
FT TSP TYPE-1 3.  
FT DOMAIN 568 625  
FT TSP TYPE-1 4.  
FT DOMAIN 548 703  
FT TSP TYPE-1 5.  
FT DOMAIN 707 759  
FT TSP TYPE-1 6.  
FT NON\_TER 766  
SQ SEQUENCE 766 AA; 85987 MW; A42613E87AE91719 CRC64;  
  
Query Match 10.4%; Score 510.5; DB 4; Length 766;  
Best Local Similarity 25.7%; Pred. NO. 1.4e-27;  
Matches 130; Conservative 57; Mismatches 188; Indels 131; Gaps 14;  
  
QY 364 CSGSEQLRACSQAQPCPEQDPDRALQCAAFNSQFMGOLYQWEPFTEVQGSQRCNCR 423  
Db 108 CEGQNIKYKTCNSNHDPPDAEDFRAQCSAYNDVQYQGHYYEWLPYRN-DPAAPCALKCH 166  
QY 424 PGFFRYVHTEKVDGTLCPQGPAPDVCVAGRCISPGCDGILGSRRRPDGGVCGGDST 483  
Db 167 AQGNLVVELAPKVLDTGRCNTDLSMDISGICQAVGCDRLGSKNAKEDNCGVCGAGDST 226

QY	484	CRVSGNLTDRGGP-LGYQKILWIPAGALRLQIAQLRPPSNYL---	ALRGPGRSIIINGN	539
Db	227	CRVSGSKSHVSEKREENVIAVPLGSRVRIIVKGPVILFIESKTLQSGKEHSFNS-	285	
QY	540	WAYDPPGSGYRAGGVFYKYNRPREGESLSABGP-----	TTQPDVYVIFQ	588
Db	286	-----PGVFFVENTVBFQ-----GSRQTFKIPGLMADFIKRYTAAKDSVQFF--	334	
QY	589	ENPGVEFYQYVSSPPPILENPTPEPPVQQLPEILRVEPPLAPAPRPTPGTLQQRVRI	648	
Db	335	-----FTQ-----PI-----	339	
QY	649	PQMAPPHPTPLGSPAAYKRWGHSACSCGKGVWRPIFLCISRSGBELDERSC---	705	
Db	340	-----SHOWRQTDFFCTVTCGGGYOLNSAECVDILKRWVFDHYCHYY	383	
QY	706	AAGARPASPPCHGTCPC-----PYWEAGWTSCSRSGCGTQHQRL	748	
Db	384	PENYKPKLKECSMDPCSSDGKEIMPYDHFQPLRWEHNPWTACVSCGGGIQRSP	443	
QY	749	QCROE-FGGGSSVPPRCGHLRPNITQSCQLRLCGHWEYGVSPWQSCVRCGRQSRQ	807	
Db	444	VCVESMHGELIQVEWKCMTAPKYMOTCNLCDFCKW-IAMWSQCTVTCGRGLRYR	502	
QY	808	VRCVGNNGDEVS-----EQEC	823	
Db	503	VLCINHRGHEVGGCNPQLKHIKEC	528	

RESULT 15

Q9SR33 PRELIMINARY; PRT: 1014 AA.

AC Q9SR33;

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE GH192189.

GN CG2131.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Y, CN BW SP;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celnik S.

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY061825; AAL27636.1;

SQ SEQUENCE 1014 AA; 112241 MW; 030A1645935D5360 CRC64;

Search completed: July 23, 2002, 21:12:41  
Job time: 475 sec

QY	307	RGRGQGGQGWGTGTPHGPRLPEPPQHPGAWPLLSNGPHASSLSLF-----	355	
Db	59	RKVGQAGGGGAGG-----GPGQW-----SSWDSWTSCTCDGGI	95	
QY	356	-----APSSPIPRCSGSEQLRACQAPCPPEQPPRALQCAAFNSQEFMGOLYQWE	407	
Db	96	MHOMRRCGSGFS-----CRGSEFYRICNMQPC-PEQQDFRSCCSAYNDVPYDGTLYKWT	150	
QY	408	PTEVQSGQRCCLNCRPRGRF-----YVHTKVDGTLIC	443	
Db	151	PHYDY--VEPCALTCRGPFAHLVEDISRETGDGNABEAHYDEQSVIQLSARVQDGTTC	208	

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 21:05:26 ; Search time 24.69 Seconds  
(without alignments)  
1375.336 Million cell updates/sec

Title: US-10-041-770-2

Perfect score: 4895

Sequence: 1 MENWGRPWLYLLLLSLPO.....PPAISCILGNHAQDTSAPFA 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query %	ID	Description
1	709	14.5	1077	1	AT10_HUMAN
2	663.5	13.6	1593	1	AT12_HUMAN
3	594	11.9	1629	1	AT59_HUMAN
4	545	11.1	1211	1	AT52_HUMAN
5	538	11.0	450	1	AT10_MOUSE
6	531	10.8	1205	1	AT52_BOVIN
7	514.5	10.5	860	1	AT56_HUMAN
8	503.5	10.3	967	1	AT51_RAT
9	501	10.2	1201	1	AT53_HUMAN
10	486.5	9.9	968	1	AT51_MOUSE
11	466	9.5	967	1	AT51_HUMAN
12	464.5	9.5	997	1	AT57_HUMAN
13	460	9.4	890	1	AT58_HUMAN
14	399.5	8.2	837	1	AT54_HUMAN
15	399.5	8.2	905	1	AT58_MOUSE
16	395	8.1	930	1	AT55_HUMAN
17	385	7.9	930	1	AT55_MOUSE
18	373	7.6	630	1	AT54_RAT
19	337.5	6.9	867	1	SSPO_BOVIN
20	277.5	5.7	2142	1	BAT2_HUMAN
21	262	5.4	660	1	YHL1_EBV
22	248	5.1	3530	1	M15_HUMAN
23	244	5.0	3511	1	M15_MOUSE
24	243.5	5.0	1183	1	DRFL_RAT
25	238.5	4.9	963	1	YQ36_CAEEL
26	235.5	4.8	620	1	EXTN_TOBAC
27	235.5	4.8	1572	1	BA12_HUMAN
28	231	4.7	1736	1	CA2B_HUMAN
29	225.5	4.6	1584	1	BA11_HUMAN
30	224.5	4.6	1185	1	DRFL_HUMAN
31	216	4.4	1453	1	CALL_CHICK
32	214	4.4	2715	1	TRX2_HUMAN
33	212.5	4.3	1496	1	CA25_HUMAN

## ALIGNMENTS

```

RESULT 1
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT: 1077 AA.
AC Q9H324;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT Thrombospondin type 1 repeats.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC similarity).
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -! SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: A7163762; BAG35563.1;
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Repolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; zn_Mtpeptdse.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC Pfam: PF01421; Repolysin; 1.
CC SMART: SM00090; TSP1; 5.
CC SMART: SM00209; TSP1; 5.
CC PROSITE: PS00215; ADAM_WEPPO; 1.
CC PROSITE: PS00092; TSP1; 2.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix.
CC -----
FT NON_TER 1
FT PROPEP <1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.

```

p53420 homo sapien  
 p04258 bos taurus  
 p81122 mus musculus  
 p02452 homo sapien  
 p12105 gallus gall  
 q98377 canis famill  
 p78359 homo sapien  
 p20508 homo sapien  
 p39060 homo sapien  
 q9nzm4 homo sapien  
 p02458 homo sapien  
 p08121 mus musculus

FT	METAL	366	366	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	367	367	BY SIMILARITY.
FT	METAL	370	370	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	376	376	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	434	520	DISINTEGRIN-LIKE.
FT	DOMAIN	578	679	CYS-RICH.
FT	DOMAIN	802	802	SPACER.
FT	DOMAIN	521	577	TSP TYPE-1 1.
FT	DOMAIN	799	860	TSP TYPE-1 2.
FT	DOMAIN	862	918	TSP TYPE-1 3.
FT	DOMAIN	922	976	TSP TYPE-1 4.
FT	DOMAIN	981	1031	TSP TYPE-1 5.
FT	CARBOHYD	64	64	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	714	714	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	769	769	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	866	866	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1077 AA;	118072 MW;	3914DE18DCBFF587 CRC64;
Query Match 14.58; Score 709; DB 1; Length 1077;				
Best Local Similarity 31.58; Pred. No. 2.4e-25;				
Matches 180; Conservative 69; Mismatches 258; Indels 64; Gaps 19;				
QY	287	QWASPVAGRRPPDPSPVPRG-RQQGQ-GPWGTGTPHGRLEPDPHPGAWLPLLSN	344	
DB	502	KGWCYKRYC-----VYFGSRPESVDGAWPWPWGDCSRTCGGVSSSRHCDSPRTIGG	557	
QY	345	GPHASLSLFPAPSPIPRCGSEBQLRACQACPPPPRALOCAFNQSFMGQLY	404	
DB	558	-----KYCLGRRHRHSNTDDCPGSGQDFREVQCCSEFDSIFPRKFY	600	
QY	405	QWEPTEVQSGORCBLNCRPGFRFYVRHTEKVDGTLCPQAPDVCVAGRCLSPGCDGI	464	
DB	601	KWKT-Y-RGGVKACSLTSLAEGFNEYTERAAAVDGTFCRPTDVCVSGECKHVGCDRV	659	
QY	465	LGSGRRPDCGVCGGDDSTCRVLSGNLTDGPGYQKILNIPAGALRLQTAQLRPSNY	524	
DB	660	LGSDLREDKRCVCGDGSACETIEGVFSPASPGAGYEDVWIPKGSVHFIQDLNLSLH	719	
QY	525	LALRPGGRSIINGWAVDPDPSYRAGTVFRYNRPREECKGSLSAEGTTPQVDVYM	584	
DB	720	LALKGDQSLLEGILPGTPQPHRLPLACTTQLRQGPQV--QSLEALGPINASLIWMV	776	
QY	585	IFQENPGVFIQY----VISPPPILENTPEPPV-----PQLQPEILRVE-PPLAPA	632	
DB	777	LARTLPAALRYFNAPLARDLSLPYSWHYAPWTKSAQCAGGSQVQAVECRNQLDSSAVA	836	
QY	633	PPARTPCTL-QROYRIPQMPAPPHPRTPLGSPAAYKRWGSHSACSASCGKGVWRPFLC	691	
DB	837	PHYCSAHSKLPRQACNTEPCPPD-----WVGNWSLCRSCDAGVRSRVSVC	885	
QY	692	IGRESGEE---LDERSACAGAPPASPCHGTGTPCPYWEAGWTCRSRSCGPGTQHRQL	748	
DB	886	QRRVSAAEKALDSDACPO-PRPPVL-EACHGTPCPPEWAALDWESECTPSGPGGLRHVV	943	
QY	749	QCRQEFGGGGSVPPRCRHLRPNITQSCQLRLC--GHWEVGSFWSQSVRCRGRQR	806	
DB	944	LCKS--ADHRATLPPAHCSAPAKPPATMNCNLRPPARWVAGE-WGECGACQGVGRQR	1000	
QY	807	QVRGVNNGDVESEQEACGPPQPPSREACD	837	
DB	1001	SVRCTSHTGQ--ASHECTEA-LRPPTQOCE	1028	
RESULT 2				
ID	AT12_HUMAN	STANDARD;	PRT;	1593 AA.
AC	P58397;			
DT	01-MAR-2002 (rel. 41, Created)			
DT	01-MAR-2002 (rel. 41, Last sequence update)			
DT	01-MAR-2002 (rel. 41, Last annotation update)			

DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).

GN ADAMTS12.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal lung;

RX MEDLINE=21264577; PubMed=11279086;

RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.; "Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";

RT J. Biol. Chem. 276:17932-17940(2001).

RL J. Biol. Chem. 276:17932-17940(2001).

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.

CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.

CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1 DOMAINS.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AJ250725; CAC20419.1; ..

CC MTM: 606184; ..

DR PROSITE; PS00215; ADAM\_MEPRO: 1.

DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.

DR PROSITE; PS00092; TSP1; 2.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix.

KW SIGNAL

FT 1 25 POTENTIAL

FT PROPEP 26 240 BY SIMILARITY.

FT CHAIN 241 1593 ADAMTS-12.

FT DOMAIN 465 544 DISINTEGRIN-LIKE.

FT DOMAIN 545 596 TSP TYPE-1 1.

FT DOMAIN 597 700 CYS-RICH.

FT DOMAIN 701 826 SPACER 1.

FT DOMAIN 827 881 TSP TYPE-1 2.

FT DOMAIN 886 943 TSP TYPE-1 3.

FT DOMAIN 947 995 TSP TYPE-1 4.

FT DOMAIN 996 1315 SPACER 2.

FT DOMAIN 1316 1364 TSP TYPE-1 5.

FT DOMAIN 1367 1423 TSP TYPE-1 6.

FT DOMAIN 1426 1471 TSP TYPE-1 7.

FT DOMAIN 1471 1471 TSP TYPE-1 8.

FT DOMAIN 302 305 POLY-GLU.

FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).

FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT\_SITE 393 393 BY SIMILARITY.

FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).

FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1593 AA;	177545 MW;	07F9F48E63BD83A3 CRC64;	
Query Match 13.68; Score 663.5; DB 1; Length 1593;					
Best Local Similarity 29.88; Pred. No. 3.6e-23;					
Matches 168; Conservative 78; Mismatches 203; Indels 115; Gaps 20;					
QY	296	GRPPFPFVPRGRQQGQGPWG-----TGGTPHGPRL--EPDPHPGAWLPLLSNGPH	347		
DB	535	GKKPE--SIPGNGR--WSPWHSRRTCGAGVQSAERLNNPFPKGGKY-----	580		
QY	348	ASSLSLAPSPPIPCGSESEQLRACSAQCPPEPPRALQCAAFNSQEFMGLQIWE	407		
DB	581	-----CTGKRKRYLCLVHPCRSEAPTFROMQSEFTVPYKNELYHWF	624		
QY	408	PTEYQSGORCELNCRGFRFYVHTKVDGTLCPQG--APDICVAGRLSPCGDIL	465		
DB	625	PI--FNPAHPCLYCRPIDGQFSEKMLDAVIDGTFCFEGGNSRVNCGICKMWGCDVEI	682		
QY	466	GSRRPDGCGVGGDDTCRLVSGNLTDRGGPLGQYKILWIPAGALRQIAQLRPSSNYL	525		
DB	683	DSNATEDRGVCLGDSGSCQTVRKFKQKGS-GYVDIGLIPKPGARDIRVMEIEGAGNFL	741		
QY	526	ALRGPG-GRSIINGNWAYDPPGYSRAGTVFVYRNPREEKGESLSAEGTTPQVDVYM	584		
DB	742	AINSEDPKXYIINGFIQWNGNYLAGTVQYDR---KGLDKLMATGPTNESVWLQ	797		
QY	585	IFQEEPNPGFYQYVSSPPPILENTPPEVPQLOPEILRVEPPLAPAPRPTGTLQR	644		
DB	798	LQVTPNGIKYEYTKDQ--LDND-----	820		
QY	645	QVRIQMPAPPHPPLPLGSPAAYKRVGH-SACSASCGKGVWRPFLCISRESGELDER	703		
DB	821	---VEQM-----YFQYGHWTCSVTCGTGIRRTAHCIKKRG-MVKAT	861		
QY	704	SCAAGARPAPSPCHGTGTPPYWEAGWTSRSCGP-GTQHQLOCRQEFGGGSSVP	762		
DB	862	FCDPETQPNRGKCKEKACAPRWAGWEACATCGPHGKKRTVLCITQTMVSDQALP	921		
QY	763	PERCHLPNPNTQSCQLR-LC-GHWEVGSFWSQSVRCGRGQRQSRQVRCVGNNGDEVSE	820		
DB	922	PTDCQHLKPKTLTSCNRDILCPDWTGVN-WSECSVSGGGVRSVTCANKH-----	974		
QY	821	QECASGPPQPSREACDMGPCTTA	844		
DB	975	DEPCDVTFRPNSEALCGLGQQCFSS	998		
RESULT 3					
AT99_HUMAN	STANDARD;	PTT;	1629 AA.		
ID	AT99_HUMAN				
AC	Q9P2N4; Q9NR29;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase				
DE	with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).				
GN	ADAMTS9 OR KIAA1312.				
OS	Homo sapiens (Human).				

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT ISOFORM).				
RC	TISSUE=Fetal;				
RX	MEDLINE=20396138; PubMed=10936055;				
RA	Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,				
RA	Maki R.A.;				
RT	"ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene				
RT	family.";				
RL	Genomics 67:343-350(2000).				
RN	[2]				
RP	SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).				
RC	TISSUE=Brain;				
RX	MEDLINE=20181126; PubMed=10718198;				
RA	Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. XVI.				
RT	The complete sequences of 150 new cDNA clones from brain which code				
RT	for large proteins in vitro.";				
RL	DNA Res. 7:65-73(2000).				
CC	-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR				
CC	MATRIX (BY SIMILARITY).				
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A				
CC	SHORT FORM: MAY BE PRODUCED BY ALTERNATIVE SPLICING.				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.				
CC	EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,				
CC	PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN				
CC	COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR				
CC	THYMUS.				
CC	-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT				
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY				
CC	SIMILARITY).				
CC	-1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY				
CC	SIMILARITY).				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.				
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.				
CC	-----				
CC	this SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
EMBL	AF261318; AAF89106.1; -				
EMBL	AB037733; BAA92550.1; -				
MIN	605421; -				
InterPro	IPR001762; Disintegrin.				
InterPro	IPR002870; Pep_M12B_propep.				
InterPro	IPR001590; Repolysin.				
InterPro	IPR000884; TSP1.				
InterPro	IPR000130; zn_Mtpeptdse.				
Pfam	PF01562; Pep_M12B_propep; 1.				
Pfam	PF01421; Repolysin; 1.				
Pfam	PF00090; tsp_1; 11.				
SMART	SM00209; TSP1; 12.				
PROSITE	PS0215; ADAM_MEPRO; 1.				
PROSITE	PS00427; DISINTEGRIN_1; FALSE_NEG.				
PROSITE	PS50092; TSP1; 9.				
PROSITE	PS00142; ZINC_PROTEASE; 1.				
Hydrolase	Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;				
Repeat	Extracellular matrix; Alternative splicing.				
SIGNAL	1 18				
PROPEP	19 287				
CHAIN	288 1629				
DOMAIN	509 587				
DOMAIN	589 642				
DOMAIN	645 752				
DOMAIN	753 880				

FT	DOMAIN	999	1053	TSP TYPE-1 2.
FT	DOMAIN	1056	1108	TSP TYPE-1 3.
FT	DOMAIN	1111	1156	TSP TYPE-1 4.
FT	DOMAIN	1184	1239	TSP TYPE-1 5.
FT	DOMAIN	1240	1295	TSP TYPE-1 6.
FT	DOMAIN	1332	1383	TSP TYPE-1 7.
FT	DOMAIN	1386	1439	TSP TYPE-1 8.
FT	DOMAIN	1445	1498	TSP TYPE-1 9.
FT	DOMAIN	1501	1554	TSP TYPE-1 10.
FT	DOMAIN	1562	1612	TSP TYPE-1 11.
FT	DOMAIN	88	96	POLY-SER.
FT	SITE	223	223	CYSTEINE SWITCH (POTENTIAL).
FT	METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	435	435	BY SIMILARITY.
FT	METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	749	749	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	840	840	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1267	1267	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1064	1072	CLVTGKGKH -> VRNEGCVFP (IN SHORT ISOFORM).
FT	VARSPLIC	1073	1629	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	367	367	F -> L (IN REF. 1).
SQ	SEQUENCE	1629 AA;	182649 MW;	C1C4CEFF58B8941F CRC64;
Query Match 11.9%; Score 584; DB 1; Length 1629;				
Best Local Similarity 26.2%; Pred. No. 1.3e-19;				
Matches 191; Conservative 73; Mismatches 235; Indels 230; Gaps 31;				
QY	316	FW--GTGTP-----HG----	PRLEPDPOHPGAWLPLLSNGPHASLSLFPASS-----	359
DB	559	PWADTEGPKGKHYGCVKEMDVPVTDGWS-----	GSWSFOTCSRTCGGG	607
QY	360	-----PIPR-----	CGSEQLRCAOAPCPPPEOPDPRALQCAAFNSQEF--MGQ	402
DB	608	IKTATRECNRPENKNGKYGVRMKFKSCNTECLKRKDRFEQCAHDFGKHFNGL	667	
QY	403	L--YQWEP-FTEVQSGQCELNCRPRGRFRVHTKRVQDGTLCQPGADICVAGRLSP	459	
DB	668	LPNVRWPKYSGILMKDKCKLFCRVAGNTAYQLRDVIDGTGPGQDNTDTCVQGLCRA	727	
QY	460	GDGTLGSRPDPGGVCGGDDSCRLVSGNLTDRGGPLGVOKILWIPAGALRLQIAQLR	519	
DB	728	GCDHVNSKARKDKCGVCGGDSNCKTVAG--TNTVHYGNTVVRIPAGATNIDVRQHS	785	
QY	520	PS-----SNYALRGPGRSINGNWAYD-EPGSYRAGGTFRYNRPREGKGSLSAE	573	
DB	786	FSGETDDNLYALSSSGKEFLNGFVVTMAKREIRIGNAVVEYS--GSETAVERINST	842	
QY	574	GPTTOP--VDVYMFQENPGVYQYVLISSPPILENTPPEPPVQLOPELLRVEPPLAP	631	
DB	843	DRIEQELLQVLVSGKLYNPDRVSFNI-----PIEDKP-----	QQFYWNHSGHWQA	889
QY	632	APRPARTGTQLQRY-----	RIPQMPAPPHPTPLGSPA-AYWKVGHSHAC	676
DB	890	CSKPCQ--GERKKLVCTRESQDLTVSDQRCDRLPQPGHTEPCGTDCDLRHWYASRSEC	947	
QY	677	SASCKGWNRPFLC--ISRSRG--EELDSRCAAGAPPAPEPCGHT-----	721	
DB	948	SAQCGLYNTLDIYCAKYSRLDGKTEKYDRDGGFCSSHPK-PSNRKCSGECNTGGWYSAW	1006	
QY	722	-----	PCPPYWEAGETWSCS	736
DB	1007	TECSKSCDGGTORRAICVNTNVDLDSKTHQEKVTIQRCSFPC-PQWKGSDWSECL	1065	
QY	737	RSCGPGTHRLQCRQETFG-----		
DB	1066	VTCGKGHKHRQWVC--QFGEDRLNRCMDPKPTKPTSMOTCOQPECASWQAGFWGQCSVTC	1123	

QY	756	GGG-----SSVPPERCGLPRPNITQSCQLRLC-----	783
DB	1124	GQGYQLRAVKCIIGTYKSVYDDNDCHNAATPTDQCELPSCHPPAAPETRRSTYSAPR	1183
QY	784	GHWEVSPSQSVRCGRQGRSRQVRCVGNNGDEVSEQCASGPPOPPPRBRACDMGPCPT	843
DB	1184	TQWRFGS-WTPSCATCGKTRMYVSCRDNGSVADSEACAT-LPRPVAKECSTVPC-G	1240
QY	844	AWPHSDWS 852	
DB	1241	QWKALDWS 1249	
RESULT 4			
ID	ATS2_HUMAN	STANDARD;	PRT: 1211 AA.
AC	O95450;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)		
DE	(Procollagen I/II amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI)		
DE	(Procollagen I/II amino-propeptide processing enzyme).		
GN	ADAMTS2 OR PCINP OR PCPNI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).		
RC	TISSUE=Skin;		
RX	MEDLINE=99347935; PubMed=10417273;		
RA	Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E., Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W., Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.; "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene."; Am. J. Hum. Genet. 65:308-317(1999).		
CC	-!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN COLLAGEN BIOSYNTHESIS.		
CC	-!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.		
CC	-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).		
CC	-!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).		
CC	-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.		
CC	-!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.		
CC	-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.		
CC	-!- PTM: THE PRECURSOR IS CLEAVED BY A FUKIN ENDOPEPTIDASE (BY SIMILARITY).		
CC	-!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS SYNDROME TYPE VIIC (EDS-VIIC), A RECESSIVELY INHERITED DISORDER CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.		
CC	-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.		
CC	-!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.		
CC	-!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```
EWBL; AJ003125; CAA05880.1; -.
MEROPS; M12.301; -.
MIM; 604539; -.
MIM; 225410; -.
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; Zn_Mipetdse.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF00030; tsp_1; 4.
SMART; SMO0209; TSP1; 4.
PROSITE; PS50215; ADAM_MEPRO; 1.
PROSITE; PSSC092; TSP1; 1.
PROSITE; PSSC042; ZINC_PROTEASE; FALSE.NEG.
PROSITE; PS00427; DISINTEGRIN_1; FALSE.NEG.
Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
Alternative splicing.
SIGNAL 1 29 POTENTIAL
PROPEP 30 253 BY SIMILARITY.
CHAIN 254 1211 ADAMTS-2.
METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
ACT_SITE 409 409 BY SIMILARITY.
METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
DOMAIN 480 560 DISINTEGRIN-LIKE.
DOMAIN 561 617 TSP TYPE-1 1.
DOMAIN 618 722 CIS-RICH.
SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).
DOMAIN 723 851 SPACER.
DOMAIN 852 911 TSP TYPE-1 2.
DOMAIN 912 974 TSP TYPE-1 3.
DOMAIN 975 1030 TSP TYPE-1 4.
DOMAIN 40 43 POLY-ALA.
DOMAIN 185 188 POLY-GLU.
CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 251 251 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 949 949 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 993 993 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 1031 1031 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 1098 1098 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 1145 1145 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 1150 1150 N-LINKED (GLCNAC. .) (POTENTIAL).
VARSPPLIC 544 566 HCFCGQHWLTPDLLARDGSWA -> PRPGVAHACYPS
TLGGQGRIIA (IN ISOFORM SNPN1).
MISSING (IN ISOFORM SNPN1).
SEQUENCE 1211 AA; 134722 MW; BCECEFF25C23CAD2D CRCS64;
BCECEFF25C23CAD2D CRCS64;
```

Very Match 11.1%; Score 545; DB 1; Length 1211;  
1st Local Similarity 26.9%; Pred. No. 5.8e-18;  
Matches 167; Conservative 71; Mismatches 247; Indels 136; Gaps 23;

```

289 WASPOVACRRDPDPFSPVPRGRGQOGPWGIGTTPHCGPRLEPPDPQHPGAWLPILLS----- 343
      | | | | |
552 WLTDPILKR-----DGSW-----GAWSPFGSCSRTC 577
      | | | | |
344 -----NGPHASLSLFPAPSPICRSGESQLRACSQAPCPPEQPPDPRALQCA 392
      | | | | |
578 GTGVKFRTRQCDNPH-----PANGRRCTGLAYDFOLCRQDCPOSLADFREEQCR 628
      | | | | |
393 AFNSQ- EFMGQLYQWEPFEVQGSQRCELNCRPGRGFYVYRHKRYQDGTGLCO- PGAPDI 450
      | | | | |
629 OWDLFYFEGDADHNLWPHEDRDKERCHLYCSGSRGTGEVYSMKRMVHDGTRCYSKDAPSL 688
      | | | | |

```

Qy	451	CVAGRCLSPCGCGDGLGSGRRPDCGCGCGDDSTCKLVSGNLITDRGGPLGYGKILWIPAGA	510
Db	689	CVRGDCRKYCGCGVIGSSGQEDKCGCGGDNHGCRVAVAGTFTSRPKKHGYIKMFEIPAGA	748
Qy	511	LRQIAQLRAPSSNYLALRG-PGGRSTIINGNAVDPPG-SYRAGGTVFPRNRPPEEGKGE	568
Db	749	RHLIIQEVATSHLAVKNLETKGKILNEENDVDASSKTFAMGVEWEY----	804
Qy	569	SLSAEGPTQPDVYMI-FOENPGVFQYVYVTSPPPILENTPEFPVQLOPILRVPEP	627
Db	805	TLQTMGLGHTITVLVIPGDTRVSLTKYMT-----HEDSLAVD-	844
Qy	628	PLAPAPPARTPGLTQROVRIPOMPAPPHPTPLGSPAAYWKRVGSHSACSACGKGVWRP	687
Db	845	-----DNNVLEED-----SVVEWALKKWSCKPCCGGGSGQFT	877
Qy	688	IFLCLISREGEELDERSCAAGRAPSPAPPCHGTPC-PPYWEAGEWTSKRSRSG-PGTQH	745
Db	878	KYGCRRRLDEKMYHRGFCALSKPAIRACNPQCSOPVVMVTEGWEPCSTCGRTGMQV	933
Qy	746	RLQLCRQEGGGGS-SVPPERCGHLPRNITQSCQLRLC-GHWEVGSWSPSCSVRCGRGQ	803
Db	938	RSVYCIQPLHDNTRVSHAKHND-ARPESSRACSRCLCPGRWRAG-FWSQCVTCGNGT	995
Qy	804	RSQVRVC-VGNNGDEYSEQEGAGPPQPPSRACDMGPC-----TTAFHSDSWSSKVS	855
Db	996	QERPVCRTADDSFGICQEE-----RPETARTCLGPCPRNISDPKSKYVYVWLSPD	1049
Qy	856	PEPPAISCILGNHAQ-DTSAPF	875
Db	1050	PDSPIRKISKGHCCQGDKSIF	1070

RESULT 5

AT10\_MOUSE

AT10\_MOUSE

STANDARD;

PRT;

450 AA.

AC

P58459;

DT

01-MAR-2002 (Rel. 41, Created)

DT

01-MAR-2002 (Rel. 41, Last sequence update)

DT

01-MAR-2002 (Rel. 41, Last annotation update)

DE

ADAMTS-10 (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).

DE

ADAMTS10.

GN

Mus musculus (Mouse).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX

NCBI\_Taxid=10090;

RP

[1]

RN

SEQUENCE FROM N.A.

RT

Seldin M.F., Apte S.S.;

RA

"A novel gene of the ADAMTS family predicts ADAMTS-10, a metalloproteinase with unique structural features and expression pattern.";

RL

Submitted. (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC

-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC

-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC

-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By similarity).

CC

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC

-1- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.

CC

-----

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions in use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC

EMBL; AF302012; RAK97226.1;

DR

----

DR PROSITE; PS00092; TSPL; 1.  
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Repeat;  
KW Extracellular matrix.  
FT NON\_TER 1  
FT DOMAIN 1 51  
FT DOMAIN 52 174  
FT DOMAIN 171 232  
FT DOMAIN 234 290  
FT DOMAIN 294 348  
FT DOMAIN 353 399  
FT CARBOHYD 86 86  
FT CARBOHYD 141 141  
FT CARBOHYD 238 238  
SQ SEQUENCE 450 AA; 48861 MW; 2377DUE2CFBFFCA CRC64;  
  
Query Match 11.0%; Score 538; DB 1; Length 450;  
Best Local Similarity 32.0%; Pred. No. 5.3e-18;  
Matches 141; Conservative 52; Mismatches 192; Indels 56; Gaps 15;  
  
QY 437 VODGTLCQPGAPDVCVAGRLSPGCDGILGSGRRPDGCGVCGDDTCRLVSGNLTDRGG 496  
DB 4 VVDGTPCFRPTVDICVSGECKHVGCDRLVGLSDREDKRCVCGDGSACETIEGVFSPALP 63  
QY 497 PLGYOKILWIPAGALRLQIAOLRPSSNYLALRGPGRSIINGNAVDPGSGYRAGGTVER 556  
DB 64 GTGYEDVWIPKSGVHIQIDNLNLSHLALKGQESLLEGLPTGPQPHRLPLAGTTFH 123  
QY 557 YNRPREGKESIAEGTTPQVDVVMFOENPGVYQVVISPP-----PPILNPT 610  
DB 124 LRQGP---DQASLEALGPINSLIIWYLAQELPALHYRF--NAPIARDALPPYSWHYA 178  
QY 611 PEPVPYQ-----LOPEILRVEPPL--APAPRPARTGTL-ORQVRIPQMPAPPHPRTPL 661  
DB 179 PWTKSAQACGSGVQVVECRNQLDSSAVPHYCSGHSKLPKRQACNTECPDP----- 233  
QY 662 GSPAAYKRVGHSACSGKGVWNPFLICISRESGEE---LDRSCAAGRAPSPASPEPC 718  
DB 234 -----WVGNWNRSCDAGVRSRVVQCRRVSAAEKALDSDACQP-PRPPVL-EAC 285  
QY 719 HCTPCPPYWEAGETWSCRSRCPGQHQLOCRQEFGGSGSVPERGHLPRNITQSC 778  
DB 286 QQPMCPPEWATLDWSECTPCGPGULHRHVLCKS--ADQRTLPFGHCLPAKPPSTWRC 343  
QY 779 QLRLC--GHWEYSGPWSQCSVRGQRQSRQVCGVNGNDEVSEQECASGPPQPPSREAC 836  
DB 344 NLRRCPPARW-VTSEWGECSQCGLGQQOQRTVRCSTHTGQ-----PSRECT 388  
QY 837 D-MGPCTTANFHSWSSKVSP 856  
DB 389 EALRPSTMQCEAKCDSVYPP 409  
  
RESULT 6  
AT52\_BOVIN STANDARD; PRT; 1205 AA.  
ID AT52\_BOVIN  
AC F79331;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and  
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2)  
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I  
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).  
GN ADAMTS2 OR NPI.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP PROPEP 29 253  
FT SIGNAL 1 28  
FT CHAIN 254 1205  
FT ADAMTS-2.

RX MEDLINE-97225960; PubMed-9122202;  
RA Collige A., Li S.W., Sieron A.L., Nussgens B.V., Prockop D.J.,  
RA Lapierre C.M.;  
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a  
RT new member of the superfamily of zinc-metalloproteinases with binding  
RT sites for cells and other matrix components.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).  
RN [2]  
RP PARTIAL SEQUENCE.  
RX MEDLINE-95348096; PubMed-7622483;  
RA Collige A., Beschip A., Samyn B., Goebels Y., Van Beemmen J.,  
RA Nussgens B.V., Lapierre C.M.;  
RT "Characterization and partial amino acid sequencing of a 107-kDa  
RT procollagen I N-proteinase purified by affinity chromatography on  
RT immobilized type XIV collagen.";  
RL J. Biol. Chem. 270:16724-16730(1995).  
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR  
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO  
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN  
CC COLLAGEN BIOSYNTHESIS.  
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
CC alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
CC at Ala-|-Gln.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO  
CC COLLAGEN TYPE XIV.  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL  
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS  
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS  
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOPARAXIS, A  
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN  
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN  
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X96389; CAA65253.1;  
DR InterPro: IPR001762; Disintegrin.  
DR InterPro: IPR002870; Pep\_M12B\_propep.  
DR InterPro: IPR001590; Reprolysin.  
DR InterPro: IPR000884; TSPL.  
DR InterPro: IPR000130; zn\_Mtpeptdse.  
DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
DR Pfam: PF01421; Reprolysin; 1.  
DR Pfam: PF00090; tsp.1; 4.  
DR SMART: SM00209; TSPL; 4.  
DR PROSITE: PS00215; ADAM\_MPEPRO; 1.  
DR PROSITE: PS00092; TSPL; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 28  
FT PROPEP 29 253  
FT CHAIN 254 1205  
FT ADAMTS-2.



FT	METAL	402	402	ZINC (CATALYTIC) (POTENTIAL).
FT	ACT_SITE	403	403	BY SIMILARITY.
FT	METAL	406	406	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	412	412	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	474	554	DISINTEGRIN-LIKE.
FT	DOMAIN	555	611	TSP TYPE-1 1.
FT	DOMAIN	612	716	CYS-RICH.
FT	DOMAIN	717	845	SPACER.
FT	DOMAIN	846	905	TSP TYPE-1 2.
FT	DOMAIN	906	968	TSP TYPE-1 3.
FT	DOMAIN	969	1024	TSP TYPE-1 4.
FT	SITE	685	687	CELL ATTACHMENT SITE (POTENTIAL).
FT	DOMAIN	31	35	POLY-ALA.
FT	DOMAIN	177	180	POLY-GLU.
FT	CARBOHYD	104	104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	245	245	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	942	942	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	943	943	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	987	987	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1092	1092	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1139	1139	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1144	1144	N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQ	SEQUENCE	1205 AA;	133887 MW;	7B5B232A45320371 CRC64;
Query Match				
Best Local Similarity 10.8%; Score 531; DB 1; Length 1205;				
Matches 164; Conservative 70; Mismatches 235; Indels 156; Gaps 22;				
QY	289	WASPVAGRRPDP	PPSVPRGCGQGPGTGTGTHGPRKLEPDHPHGWMLPLLS-----	343
Db	546	WLTPDILKR-----	DGNW-----	GAWSFGSCSRTC 571
QY	344	-----	NGPHASSLWSLAPSPICPGSGEQLRACSOAPCEQPDPRALQCA	392
Db	572	GTGVKPTRCDNPH-----	PANGRTCSGLAYDFQLCNSQCPDADLADFREEQCR	622
QY	393	AFNSQ-EFMQLQWEPFTEVQSRCENCRPRGRFRFYRHTKQVDTLQO-PGAPDI	450	
Db	623	QWDLYFEHGAQHHLPHEDHRAKCHLYCESKETGEVVMKRVHDTRGTSYKDAFSL	682	
QY	451	CVAGRCLSPCDGILSGRRPDCGVCVGGDDSTCLVSLGNLDRGGPLCYQKTLWPAGA	510	
Db	683	CVRGDCRKVCDDGIVGSSKOEDKCGVCGGDNHCKVKGTFSSPKKLGIYIKMFELPAGA	742	
QY	511	LRLIQALRPSNNLALRG-PGGRSIINGNVAVDPPG-SYRAGTVFRYNRPREEGKE	568	
Db	743	RHLIQAEDTSHHLAVKNLETKCKFILNEENDVDPSKTFIANGVEWEX-----	RDGEGRE 798	
QY	569	SLSAEGTTPQVDVYIMFQ-EENPGVFYQVVISPP-----	PILENPTPEPPVPLQPE	621
Db	799	TLQTMGLHGTITVLVIPEGDARISLTYKMIHEDSLNVDNNVLED-----		845
QY	622	ILRVEPLAPAPRPTGLTQVRIPQMPAPPHPRTPLGSPAAYWKRVGHSACSASCG	691	
Db	846	-----	DSVGEWALKWSPCKPCG	865
QY	682	KGVWRPFLICISRESGELDERSCAAGARPASPCHGTPC-PPYWEAGWTSCSRSCG	740	
Db	866	GGSQFTYKGRRLDHKWHVGFCDVSFKPKAIRRICNPQECSPQVYVTVGEWPCSRSCG	925	
QY	741	PGTHQLQCRQFEGGGS-SYPPERCGLHPRNTQSCQLRLC-GHWEVGSFWSQCVS	797	
Db	926	RTGMQVRSVRCVQLPHNHTFVHTKICND-ARPEGRACNRELCPGRWAGS-WSQCSV	983	
QY	798	RCGQRGRSQRVC-VGNNGDEVSEQACSGAPPPPSREACDMGPC-----	TTAFHSD 849	
Db	984	TCGNGTQEPVLCRTADDSFGVCREE-----	RPTARICRLGCPRNTSDPSKKSYYVQ	1037
QY	850	WSSKVSPEPPAICILGNAQDTSA	874	
Db	1038	WLSRPDPNSP-----	VOETSS 1053	

RESULT 7  
AT56\_HUMAN STANDARD; PRI; 860 AA.  
ID Q9UKP5;  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-6 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).  
GN ADAMTS6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=93395124; PubMed=10464288;  
Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;  
"ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of Zinc Metalloproteases.";  
J. Biol. Chem. 274:25955-25963(1999).  
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).  
CC -|- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY DETECTABLE IN A NUMBER OF OTHER TISSUES.  
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -|- FM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: AF140674; AAD56357.1; -  
HSP; P34179; 11AG.  
MIM: 605008;  
InterPro: IPR001762; Disintegrin.  
InterPro: IPR002870; Pep\_M12B\_propep.  
InterPro: IPR001590; Repolysin.  
InterPro: IPR000884; TSP1.  
InterPro: IPR000130; Zn.Mtpeptidse.  
Pfam: PF01562; Rep\_M12B\_propep; 1.  
Pfam: PF01421; Repolysin; 1.  
Pfam: PF00090; tsp\_1; 1.  
SMART: SM00209; TSP1; 1.  
PROSITE: PS00215; ADAM\_MEPRO; 1.  
PROSITE: PS00142; ZINC\_PROTEASE; 1.  
PROSITE: PS00092; TSP1; 1.  
PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
Repeat; Extracellular matrix.  
POTENTIAL.  
SIGNAL 1 21  
FT PROPEP 22 244 BY SIMILARITY.  
FT CHAIN 245 860 ADAMTS-6.  
FT METAL 403 403 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 404 404 BY SIMILARITY.  
FT METAL 407 407 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 453 509 DISINTEGRIN-LIKE.  
FT DOMAIN 510 566 TSP TYPE-1 1.  
FT DOMAIN 567 668 CYS-RICH.  
FT DOMAIN 669 795 SPACER.

```

FT DOMAIN 796 852 TSP TYPE-1 2.
FT FT 68 71 POLY-ARG.
FT FT 662 665 POLY-GLY.
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 850 AA; 97098 MW; E57213015DECB2C5 CRC64;

Query Match 10.58; Score 514.5; DB 1; Length 860;
Best Local Similarity 31.28; Pred. No. 1e-16;
Matches 125; Conservative 44; Mismatches 147; Indels 85; Gaps 9;

QY 304 SVPRGRGQGGPGTGTGTPHGRLEDDPQHPGAWLPPLLSNGPHASSIW----- 352
DB 475 SIPAAEGTLQ-----TGNIEKGWCYQDCVFGTWPQSIDGGWGPWSLWGCSTCGGV 530
QY 353 --SIFAPSPIPR-----CSGESQLRACSOAPCPPEPPDPRALQCAAFNSQEFMGQLYQ 405
DB 531 SSSLRHCDSPAPSGGKYCLGERKRYKSCNTDPCPLGSRDFRERQCADFDMNPRGRKYN 590
QY 406 WEPFTEVGSORCELNCRPRFVRHTEKVDQGTLCQCAPDICVAGRCLSPGCDGIL 465
DB 591 WKPYTG-GGVKPCALNCAEINGFYTERAPAVIDGTQCNADSLDICECKHVCNDIL 649
QY 466 GSGRRPDGCGVGGDDSTCLVSNLTD---RGGLPYQKILWIPAGALRIQIAQLRPS 522
DB 650 GSDAREDCRCVGGGSGTCDATGFEFNDLPRG---GYMEVVQIPRGSHVIEVREAVMSK 706
QY 523 NYLALRGPGRSINGNWAYDPPGYSRAGGVFRYRNPREEGKSGESLSASGPTQPDV 582
DB 707 NYIALKSGDDYYINGANTIDWPKRFVDVAGTAFYKRPTEP---ESLEALGPTSENILV 763
QY 583 YMIFOEENPGVYQYVVISPPPILENTPPEPPVQLOPEILLRVEPPLAPAPARTPQT 642
DB 764 WVLLOEQNLGIRYKENV-----WNHPWSECSATCAGG 811
QY 643 QRQVRIPQMPAPPHPRTPLGSPAYWKRVGHSACSASCGKG 683
DB 789 DNEVGFT-----WNHPWSECSATCAGG 811

RESULT 8
ATSL RAT STANDARD; PRT; 967 AA.
AC Q9W001; Q9ERIL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RL thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luethi M., Hoessli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in

```

```

RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-|-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF149118; AAD34012.1; -.
CC EMBL: AF304446; AAG29823.1; -.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Repolysin.
CC InterPro: IPR000884; TSP1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC Pfam: PF01421; Repolysin; 1.
CC Pfam: PF00090; tsp.1; 3.
CC SMART: SM00209; TSP1; 3.
CC PROSITE: PS0215; ADAM_MEPRO; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS00092; TSP1; 2.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC SIGNAL 1 54 POTENTIAL.
CC PROPEP 55 252 BY SIMILARITY.
CC CHAIN 253 967 ADAMTS-1
CC SITE 205 205 CYSTEINE SWITCH (POTENTIAL).
CC METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
CC ACT_SITE 402 402 BY SIMILARITY.
CC METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
CC METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
CC DOMAIN 476 558 DISINTEGRIN-LIKE.
CC DOMAIN 559 615 TSP TYPE-1 1.
CC DOMAIN 616 724 CYS-RICH.
CC DOMAIN 725 857 SPACER.
CC DOMAIN 858 907 TSP TYPE-1 2.
CC DOMAIN 908 967 TSP TYPE-1 3.
CC DOMAIN 194 198 POLY-ARG.
CC CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 21 21 I -> V (IN REF. 2).
CC CONFLICT 26 31 KFRSQ -> RSRGSL (IN REF. 2).
CC CONFLICT 49 49 V -> A (IN REF. 2).
CC CONFLICT 72 72 R -> P (IN REF. 2).
CC CONFLICT 79 79 L -> TR (IN REF. 2).

```

FT CONFLICT 249 249 R -> G (IN REF. 2).  
FT CONFLICT 262 265 TMLV -> NLLK (IN REF. 2).  
FT CONFLICT 607 607 S -> F (IN REF. 2).  
FT CONFLICT 936 936 L -> V (IN REF. 2).  
FT CONFLICT 962 962 I -> T (IN REF. 2).  
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCB4CF CRC64;

Query Match  
Best Local Similarity 10.3%; Score 503.5; DB 1; Length 967;  
Matches 197; Conservative 99; Mismatches 331; Indels 315; Gaps 34;

QY 74 PTVQLHPSLPPR-----PPR-----HPEALLPRQGGPRPQTSPE 109  
Db 168 PAVATERLVPAEPKESIAPRPHILRRRRSGSGAKGVWDEETLPTSNSGRESQNTPD 227  
QY 110 TPLIYRTQSGRGGPLRGPAASHLGRRETEQIEAARRRLRDLKPKMGFGVRVPPFALPLH 169  
Db 228 QWPLRNPPTQAGRP-TGPGSIRKREKVSSPRVYETMLVADQSMADPHSGSLKHYLLTLF 286  
QY 170 -----RNRHPRSPRSELSL-----SSRGEAIPSPTPRAEPFSA-----NGSPQT 212  
Db 287 SVAARFYKHPSI--RNSISLVVVKILVYEQKGPVTSNAALTLRNFCSWQKOHNSP-S 343  
QY 213 ELPTTELSVHTPSPAQLSPETAQT-----EVAPRTRPAPLRHHPRAQAASGTEPPSP 265  
Db 344 DRDPEHYDTAILFTRQDLGSGHTCDTLGMADVGVCDPSRSCSVIEDDGLQAAPT-----T 399  
QY 266 THSLGE-----GGFFRASPQPRPSSQGW----- 290  
Db 400 AHELGHVNMHDDAKHCASTNGVSGDUSHLMASLSSLDHSQSPSCSAVYMTSFLDNHG 459  
QY 291 -----SPQVAGRPPDPSPVPRGGQGGQPGW-----TGG- 321  
Db 460 GBCILMDKPNKIKPLSDPLGLTYLDANROCFTEGSESTHCPDAASTCTSLWCTGTSGLL 519  
QY 322 -----THGPRLEPDPHPQAW-----LPLLSNGPHASSLWSLAP----- 357  
Db 520 VCQTKHFWADTSCGEGKWCVSGKCVNKTDMKFATFVHGS--WGWPMPGDCSRICGG 577  
QY 358 -----SSPIPR-----CGSEQLRACSOAPPEQDPD-RALQCAAFN--SQEEM 400  
Db 578 GVOYIMRECDNPVNGKGYCEGKRVYRSCNIEDCPDNNGKTFRFEOCEAHNEFSKASE 637  
QY 401 GQ--LYQWEP-FTEVQGSORCLNCRPRGFRFVYHTKVDGILCQGPADICVAGRL 457  
Db 638 GNEPTVETWPKYAGVSPKDRCKLTCEAKGIGYFFVLQPKVYVDGTPCSPDSTSVCCVQGGCV 697  
QY 458 SPGCDGILGSRPRPDGCGVGGDDTCLVSGNLTRDGGPLGYQKILWIPAGALRLQIAQ 517  
Db 698 KAGCDRIIDSKKFKDKGCVGGNGSTCKKISGTVISTRP--GYHDIIVIPAGATNIEVKH 755  
QY 518 LRP-----SSNYLALRPGGRSIIINGM---AVDPGYSRAGGTVFYRNRPRREGKGES 569  
Db 756 RNPGRSNGSFLAIRAADGTVILANGNETLSTLEQDLAYK--GTVLRY-----GSS 805  
QY 570 LSAE-----GPTQP--VDVYMFQEEHPGVYQVVISPPPILENPTEPPVPLQPEI 622  
Db 806 AALERIRSPLEKEPTIOLVMVGRALRPIKITYTFMK----- 844  
QY 623 LRVEPPLAPAPARTPGTLQVRQPOMPAPPHPRTPLGSPAAYKRVGHSACSACGK 682  
Db 845 ----- 844  
QY 683 GVMRPIFLCISRESGELDERSCAAGRAPSPERCHGTPCPYWEAGBWTSCSRSCGPG 742  
Db 845 -----KTEPNAIPTSEWVIEGWGCSKTCGSG 873  
QY 743 TOHRLQCRQCFEGGSSVPPERCGLHPRNITQSQLRLCGHWEVGSWQSVRCRGK 802  
Db 874 WQRRVVECDINGH-----PASECAKEVPASTRPCADLPCPRWQVGD--WSPCKTCKGK 927  
QY 803 QRSRQVRVGNNGNDEVSEECASGPPQPPSR--EACDMGPT 842

Db 928 YKRLTKLCLSHDGGVLSNESC--DPLKKPKHYIDFCILTQCS 967

RESULT 9  
ATSS3\_HUMAN  
ID ATSS3\_HUMAN STANDARD; PRT; 1201 AA.  
AC O15072;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Fragment).  
GN ADAMTS3 OR KIAA0366.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97349984; PubMed=9205841;  
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT Prediction of the coding sequences of unidentified human genes. VII.  
RT The complete sequences of 100 new cDNA clones from brain which can  
code for large proteins in vitro.;  
RL DNA Res. 4:141-150(1997).  
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
MATRIX (BY SIMILARITY).  
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUFIN ENDOPEPTIDASE (BY  
SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB002364; BAA20821.1; -  
CC MEROPS; M12.220; -  
CC MIM; 605011; -  
CC InterPro; IPR001762; Disintegrin.  
CC InterPro; IPR002870; Pep\_M12B\_propep.  
CC InterPro; IPR001590; Repolysin.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR000130; zn\_Mipeptidse.  
CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC Pfam; PF01421; Repolysin; 1.  
CC Pfam; PF00090; tsp.1; 4.  
CC SMART; SM00209; TSP1; 4.  
CC PROSITE; PS00215; ADAM\_MEROP; 1.  
CC PROSITE; PS00092; TSP1; 2.  
CC PROSITE; PS00142; ZINC\_PROTEASE; FALSE-NEG.  
CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE-NEG.  
CC Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT NON\_TER 1  
FT SIGNAL <1 ? POTENTIAL.  
FT PROPEP ? 245 BY SIMILARITY.  
FT CHAIN 246 1201 ADAMTS-3.  
FT METAL 394 394 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 395 395 BY SIMILARITY.  
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).  
FT

FT	METAL	404	404	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	466	546	DISINTEGRIN-LIKE.
FT	DOMAIN	547	603	TSP TYPE-1 1.
FT	DOMAIN	604	708	CYS-RICH.
FT	DOMAIN	709	840	SPACER.
FT	DOMAIN	841	898	TSP TYPE-1 2.
FT	DOMAIN	899	961	TSP TYPE-1 3.
FT	DOMAIN	962	1013	TSP TYPE-1 4.
FT	DOMAIN	242	245	POLY-ARG.
FT	CARBOHYD	79	79	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	115	115	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	238	238	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	810	810	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	938	938	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1201 AA:	135113 MW: D54EA92BD506A3AA	CRG64;
Query Match				
Best Local Similarity		10.2%;	Score 501;	DB 1; Length 1201;
Matches 187;		Conservative 68;	Mismatches 272;	Indels 166; Gaps 31;
Qy	227	QAEPLSPETAQTEV-APRTAPLRHRAQASGTEPPSPTHS----	LGEGGFFRASPOP	281
Db	407	QGNRCGETANGVMAPLVQAAPHRYH-WSRSGQELKRYIHSYDCLLDD-----		455
Qy	282	RRPSSQGWASQVAGRRPDPFSPVPRGSGQGGPWGTC-----		320
Db	456	--PFDHWP-----KLPE-LPGINTSMDEQCREFDGVGYKMTAFRFDPCQIWCWHP		506
Qy	321	-----OTPGHLEPDPQHPGAWL-----	PLLSNPHASSLWSLPAPS----	359
Db	507	DNPFYCKTKKGLDTECAAGKWCYKGMKNANQKQDGNWS--WTKFGSCSRTCG		564
Qy	360	-----PTP-----RCGSEQLRACSAQCPPEQDPDRALQCAAFNSO-ERWG		401
Db	565	TGVFRTRQCNPNPIMGDQCGVNFVQLCNTBECQKHFEFRAQCCQQRNSHFETQN		624
Qy	402	QLYQWEPTEVQSGRCELKNCRRP--GFRFYVHTKVKVQDGLQOPGAP-DICVAGRCL		458
Db	625	TKHMLPYEHPDPKPKCHLYCQSKETGDVAYMK--QLVHDGTHCSYKDPYSICVYRGEVK		682
Qy	459	PGCDGILGSRRRDGGCGVCGDSDTCLVSGNLTDGGLPLGYQKILWIPAGALRLQIAQL		518
Db	683	VGCDKEIGSKNVEDKGVCGDSDSHCHRTVGTFTPRKLGKLMFDPDPPGARHVLQED		742
Qy	519	RPSSNYLALRGP-GRSINGNNAVDPGSGYRAGGTVFYRNPPEEGKGESLSAEGPTT		577
Db	743	EASPHILAIKNAQTHGYILNGKEEAESRTFIDLGVEDWYN----IEDDIESLHTDGLPLH		798
Qy	578	QPVDMYIFQEN--PGVYQYVI--SSPPPILENTPPEPPVLPQLPETLRVEPPLAPAP		633
Db	799	DPVILVLIPOENDTRSLTYKIIHEDSVPTNSNV-----IQEELDTFE-----		844
Qy	634	RPARTPOTLRQVRIPOMAPPHPRTPLGSPAAYKRWGHSACSACSGKGVWRPFLCIS		693
Db	845	-----WALKSWSQVSRKPCGGGFOYTKYGC-R		869
Qy	694	RESGEELDEERS-CAAGARPPASPEPCHTGCP-PLYEAEWTSKCSRSCG-PCOTQRQLQC		750
Db	870	RKSDNKVHRSFCEANKPKPIRMKNIQECTHPLWVAEWEHCTKCGSSGYQLTVRC		929
Qy	751	QRE-FGGGSGSVPPRGHPLPRNITQSCQLRLC-GHNEVGSFPWSCSVRCGQRGRSQV		808
Db	930	LQPLLDGNTNRSVHSKYCMG-DRPESRRPCNRVPCPAQWKTG-PWSECSVTGEGTEGVQV		987
Qy	809	RCVGNNGDSEQBCASGPPQPPSRACDMGPC		841
Db	988	LC--RAGDHCDGE-----KPESVRACQLPPC		1011
RESULT 10				

ATSL_MOUSE	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
ID	P97857	054768;		
AC	30-MAY-2000	(Rel. 39, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DE	ADAMTS-1 precursor (EC 3.4.24.-)	(A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).		
GN	ADAMTS1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=129/SVJ;			
RA	MEDLINE=98110583; PubMed=9441751;			
RT	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RL	"The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=895297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,			
RT	Matsushima K.;			
RL	"Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene.";			
RL	J. Biol. Chem. 272:556-562(1997).			
RN	[3]			
RX	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RA	MEDLINE=99303657; PubMed=10373500;			
RT	Kuno K., Terashima Y., Matsushima K.;			
RL	"ADAMTS-1 is an active metalloproteinase associated with the extracellular matrix.";			
RL	J. Biol. Chem. 274:18821-18826(1999).			
RN	[4]			
RP	FUNCTION.			
RX	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,			
RT	Ohno H., Matsushima K.;			
RL	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RL	FEBS Lett. 478:241-245(2000).			
RN	[5]			
RP	FUNCTION, AND INDUCTION.			
RX	MEDLINE=2043757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,			
RT	Richards J.S.;			
RL	"Progestosterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases.";			
CC	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).			
CC	INVOLVED IN ITS TURNOVER, HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY), ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY SIMILARITY).			
CC	- - CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1691-GLU- -LEU-1692 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.			
CC	- - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	- - SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX.			
CC	- - INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA CELLS OF PREOVULATORY FOLLICLES.			
CC	- - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	- - PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.			
CC	- - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	- - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.			
CC	- - SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.			



RX MEDLINE-20181126; PubMed-10718198;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.,  
RT "Prediction of the coding sequences of unidentified human genes. XVI.  
RT the complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20289799; PubMed-10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordsiek G., Horstischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dargatz E.,  
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehrach H., Reinhardt R., Raspo M.-L., 21.";  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
RN [6]  
RP SEQUENCE OF 418-967 FROM N.A.  
RC TISSUE=Melanoma;  
RA Blum H., Bauersachs S., Meves H.-W., Weill B., Wiemann S.;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOLYCAN, AND MAY BE  
CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR  
CC ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH  
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER  
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.  
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-|-LEU-1939  
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF170084; AAF15317.1; -;  
DR EMBL; AF060152; RAD48080.1; ALT\_INIT.  
DR EMBL; AF207664; BAF23772.1; -;  
DR EMBL; AB037767; BAA92584.1; ALT\_INIT.  
DR EMBL; AP001697; BAA95502.1; -;  
DR EMBL; AL162080; CAB82413.1; -;  
DR MIM; 605174; -;  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR002870; Pep\_M12B\_propep.  
DR InterPro; IPR001590; Repolysin.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF01462; Pep\_M12B\_propep; 1.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF00090; tsp.1; 3.  
DR SMART; SM00209; TSP1; 3.  
DR PROSITE; PS02015; ADAM\_MEPRO; 1; FALSE\_NEG.  
DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
DR PROSITE; PS50092; TSP1; 2.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 49 POTENTIAL.  
FT PROPEP 50 252 BY SIMILARITY.  
FT CHAIN 253 967 ADAMS-1.  
FT SITE 198 198 CYSTEINE SWITCH (POTENTIAL).  
FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 402 402 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 476 559 DISINTEGRIN-LIKE.  
FT DOMAIN 560 616 TSP TYPE-1 1.  
FT DOMAIN 617 724 CYS-RICH.  
FT DOMAIN 725 849 SPACER.  
FT DOMAIN 850 908 TSP TYPE-1 2.  
FT DOMAIN 909 967 TSP TYPE-1 3.  
FT DOMAIN 843 846 POLY-LYS.  
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 227 227 P -> A (IN REF. 4 AND 5).  
FT CONFLICT 468 468 Q -> H (IN REF. 1).  
FT CONFLICT 561 561 S -> N (IN REF. 1).  
SQ SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;  
  
Query Match 9.58; Score 466; DB 1; Length 967;  
Best Local Similarity 22.98; Pred. No. 1.6e-14;  
Matches 136; Conservative 71; Mismatches 179; Indels 208; Gaps 22;  
  
Qy 318 GTGG-----TPHGPRLEPDPQHFGAWLPLLSNGP-----HASSLSLWLFAP 357  
Db 513 GTSGLVLCQKHPWADGTCGEGKW--CINGKCVNKTDRKHDFPFHGS--WGMWGP 567  
  
Qy 358 -----SSPIPR-----CSGESEQLRACSOAPCEPQDP-RALOCA 392  
Db 568 WDCSRCTGGGVQVYTMRECDNPKNGKGYCEGRVRYRSCNLEDCCPDNNGKTRFESQCE 627  
  
Qy 393 AFNSQEF-----MGQLYQWEP-FTEVQGSQRCNELNCRPRGFYRYHRTKVDGTLCQP 445  
Db 628 AHN--EFSKASFGSGPAVEWIPKYAGVSPKDRCLICQAKGIGYFFVLQPKVVDGTPCSP 685  
  
Qy 446 GAPDICVAGRLSPGCDGILGSGRRPDCGCGGDDSTCRVLSGNLDRGGPLGYQKILW 505  
Db 686 DSTSVCGVQCCKAGCDRIIDSKKKFKDCKGCGGNGSTCKKISGVT--SAKPGYHDIIT 743  
  
Qy 506 IPAGALRLQIAQ-----LRPSSNYLALRPGGSGIINGNW---AVDPGSGYRAGTGVFVRY 557  
Db 744 IPTGATNIEVKQRQGRSNNNGSFLAKAAGDTYILNGDYTLSTLEQDINXK--GVVLYR 801  
  
Qy 558 NRPPRECKGESLSAE-----GPTQP--VDVYMFQENPGVYQYVVISPPPILENPT 610  
Db 802 S-----GSSAALERIRSFSLKEPLTIQVLTGVNLRPKIKYTYFVKK----- 844  
  
Qy 611 PEPVPVQLQPEILRVEPLAPAPRPPARTPGTLQQRVRIQPMAPPHPRTPLGSPAAYWKR 670  
Db 845 ----- 844  
  
Qy 671 VGHSACSACGKGVWRPIFLCISRESGEELDERSCAAGARPPASPECHGTFCPPYWEAG 730  
Db 845 -----KKESEFNAIPTFSAWVIE 861  
  
Qy 731 EWTSCSRSCPGTQHQLOCRQETFGGGSSVPPERCGLHPRNLTQSCQLRCLCGHWEVGS 790  
Db 862 EWGECSSKCELGWRRLVECDRNG-----QPASECAKEVKAPSTRPCADHPCFQWLIGE 916  
  
Qy 791 FWSQCSVRCGRQSRQVRQVNGNNGDEVSEQECASGPPQPPFSR--EACDMGSPCT 842  
Db 917 -WSSCSKTCGKGYKRSRLKCLSHDGGVLHSESC--DPLKKPKKHFDFTMAECS 967  
  
RESULT 12



RESULT 13  
ID AT58\_HUMAN STANDARD; PRT; 890 AA.  
AC Q9UP79; Q9NZS0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DT ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)  
DE (METH-8).  
OS ADAMTS OR METH2.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI-TaxID-9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=99367466; PubMed=10438512;  
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,  
RA Lombardo M., Iruela-Arispe M.L.;  
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new  
RT family of proteins with angio-inhibitory activity.";  
RL J. Biol. Chem. 274:23349-23357(1999).  
RN [2]  
RN SEQUENCE OF 195-440 FROM N.A.  
RX MEDLINE=20079168; PubMed=10610729;  
RA Georgiadis K.E., Hirohata S., Seidlin M.F., Apté S.S.;  
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on  
RT mouse chromosome 9 and human chromosome 11.";  
RL Genomics 62:312-315(1999).  
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
CC MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER  
CC KIDNEY.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF060153; AAF48081.1; -  
CC EMBL; AF175283; AAF25806.1; -  
CC HSSP; P34179; 1IAG.  
CC MIM; 605175; -  
CC InterPro; IPR001762; Disintegrin.  
CC InterPro; IPR002870; Pep\_M12B\_propep.  
CC InterPro; IPR001590; Repolysin.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR000130; Zn\_MTPeptidse.  
CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC Pfam; PF01421; Repolysin; 1.  
CC Pfam; PF00090; tsp\_1; 2.  
CC SMART; SM00209; TSP1; 2.  
CC PROSITE; PS0215; ADAM\_MEPRO; 1.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC PROSITE; PS50092; TSP1; 1.  
CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 27  
FT PROPEP 28 214 BY SIMILARITY.  
FT CHAIN 215 890  
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 365 365 BY SIMILARITY.  
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DOMAIN 439 526 DISINTEGRIN-LIKE.  
FT DOMAIN 527 583 TSP TYPE-1 1.  
FT DOMAIN 584 690 CYS-RICH.  
FT DOMAIN 691 832 SPACER.  
FT DOMAIN 833 890 TSP TYPE-1 2.  
FT CARBOHYD 202 205 POLY-PRO.  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 195 195 E -> R (IN REF. 2).  
FT CONFLICT 413 440 YLTSLDGGHGGCLLDADGAALPLTGL -> FSGHLLQGW  
FT IFHKYLCKVSELKDLMP (IN REF. 2).  
SQ SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;

Query Match 9.4%; Score 460; DB 1; Length 890;  
Best Local Similarity 24.9%; Pred. No. 2.8e-14;  
Matches 169; Conservative 76; Mismatches 225; Indels 208; Gaps 28;

QY 150 DPKPK-----GMGCGYVPEFALPLHRRNRHPRSPR-----SEL-----SLISSRGEA 193  
DB 375 DSKKPCTRLFGPMGKHVMAPLVHNLQTLNWPSPCSAMYTELLDGGHGGCLLDADG-AA 433  
QY 194 IPSPTPRAEPFSANGSPQTELPTELTVHPTSPQAEPL-----SPETAQTEVAPTRP 246  
DB 434 LPLPTGL-----PGMALYQDQOCROIFGDFRHCNPNTSAQDVC-----473  
QY 247 APLRHHRAQASGTEPSPHSLGEGFFFRASQPRPS-----SQGWASPOVAGRRDPDP 303  
DB 474 AQLMCH-----TDGAEPLCHTKN-----GSLPWADTQPGHLCSEGSCLPEEVEKRP-- 524  
QY 304 SVPRGRQQCGGPGWT-----GGT--PHGPRLEPDPQHPGAWLPLLSNGPHASSLSLF 355  
DB 525 -VVDG-GWAPWPGWEGCSRTCCGGGVQFVSHRECKDPEPONGRY-----565  
QY 356 APSPIPRCSGESEQLRACQACPPPEQPPRALCAAFNSQEFM---GOLYQWEP-FTE 411  
DB 566 -----CLGRRAKYQSCHTCECPDCKSFQEQCKEYKAYNAYTMDGNLQWPKYAG 617  
QY 412 VQGSORCELCNCRPFRFYVRHTEKVDGTLQCPADPICVAGRCISPGCDGILGSGRRP 471  
DB 618 VSPDRCKLFCRAGRSEFKVFEAKVIDGTLQGPETLAICVRCQCVKAGCDHVVDSPKRL 677  
QY 472 DGCVCVGGDDSTCLVSGNLTDRGGPLGYOKILWIPAGALRLQIAQ-----LRPSSNYLA 526  
DB 678 DKCVCGCGKGNCKVKVSGSLTPTN--YGYNDIVTIPAGATNIDVKORSHPGVQNDGNLA 735  
QY 527 LRGPGRSILINGWAVDP-FGSYRAGTGVFRNRPPEEGKGSLSAEGTTPQVDVYMI 585  
DB 736 LKTADGGYLLNGNLAIASIAEQDILVKGILKY-----SGSIATLERLSQF-- 780  
QY 586 FQENPGVFYQYVISSPPPILENTPTEPPYVQLOPEILRVEPLAPAPRPTGTLQOR 645  
DB 781 -----RPLPLETLVQL-----791  
QY 646 VRIPQMPAPHPPTPLGSPAAYWKVCHSACSASCGKGVWRPIFLCISRESBELDERSC 705  
DB 792 LTPVEGVFPP-----KVKIT-----FFVNDVDFSMQSK---ERAT 825  
QY 706 AAGARPPASPEPCHTGTPCPYWEAGWTSCRSRCPGTQHRQLQCRQFEGGGSSVPPER 755  
DB 826 TNIQLPLLHAQ-----WVLGDWSECSCTGAGWQRRTVECRDPSQASAT-----870



```

QY 766 CGHLPRPNITQSCQLRLC 783
DB 871 CNKALPEDAKPCESQLC 888

RESULT 14
ATSA4_HUMAN
ID ATSA4_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1)
DE ADAMTS4 OR KIAA0688.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403980; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. X.
RA The complete sequences of 100 new cDNA clones from brain which can
RA code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RA "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RA family of proteases.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -|- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -|- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX (BY SIMILARITY).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -|- INDUCTION: BY INTERLEUKIN-1.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

```

-|- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN  
 -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AB014588; BAA31663.1; -  
 EMBL; AF148213; AAD41494.1; -  
 EMBL; AY044847; AAL02262.1; -  
 HSP; P34179; ILAG;  
 MEROPS; M12.221; -  
 MLM; G03876; -  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR001590; Reprolysin.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR000130; Zn\_MTpeptdse.  
 DR Pfam; PF01421; Reprolysin; 1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS0215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS0032; TSP1; 1.  
 DR PROSITE; PS00427; DISINTEGRIN 1; FALSE NEG.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Extracellular matrix; 1  
 FT SIGNAL 1  
 FT PROPEP 52 212  
 FT CHAIN 213 837  
 FT SITE 194 194  
 FT METAL 361 361  
 FT ACT\_SITE 362 362  
 FT METAL 365 365  
 FT METAL 371 371  
 FT DOMAIN 437 519  
 FT DOMAIN 520 576  
 FT DOMAIN 577 685  
 FT DOMAIN 686 837  
 FT DOMAIN 247 252  
 FT CARBOHYD 68 68  
 FT CONFLICT 77 77  
 FT CONFLICT 626 626  
 FT CONFLICT 682 682  
 FT SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;

Query Match 8.28; Score 399.5; DB 1; Length 837;  
 Best Local Similarity 25.38; Pred. No. 1.4e-11;  
 Matches 168; Conservative 70; Mismatches 264; Indels 161; Gaps 35;

QY 61 GGVQRRSRTQCLTVQL--HPSLPPLPPRRHPHALLPR-----GQPR-PQTSF---E 109  
 DB 235 GAGLKRYLLTMAAAAKAFKHPSI-----RNPVSLVTLRLVILSGEGEPQVGPSSAAQ 287  
 QY 110 TLPLYRTQSRGGLRGCPASHLG-----RETFQETRAARRSLRDPKRGMEGYGRVP 163  
 DB 288 TLRSFCAMQRLNTPEDSDPHFTAILFTQDLCGVSTC-----DTL--GMADVGV- 338  
 QY 164 FALPLHRRNRPRSPRSELSISRG--EEAIPSPTPRAEPFSA---NGSPQTEL-PPTE 218  
 DB 339 -----CDPARSCAIVEDDGLQSAFTAHELGHVFNLMHDSKPCISLNGPLS 385  
 QY 219 LSVHTPPSQAPLSPETAQTEVATRPAPLRHPHRAQASG-----TEPPSPH----- 267  
 DB 386 TSRHYMAPVMAHVDPEPWS-----PCSAFTFDLDNGYGHCLLDKPEAPLHLPVTF 438  
 QY 268 -----SLGEGFFRASPPRPSQGWASPOVAGRRPDPFPSPVPRGQQGQP 316  
 DB 439 PGKDYDADRCQLIFGPDSPRHCPLPPCAALWCSGHLNG-----HAMCQTKHSP 488

QY 317 WGTGTPHGR-----LEPD-----PQHGAWLPLLSNGPHASSLSLWLPAPS-- 359  
Db 489 W-ADGTPCGPAQACMGGRCLHWDQLQDFNIPQ-AGGW-----GP-----WGPDGCSRT 535  
QY 360 -----PIPR-----CSGSEQLRACSQAPCPPEQP-DPRALQCAAFNS--- 396  
Db 536 CGGVQVFSRDCRTPVRNGKYGEGRTFRSCNTEDCPTGSALTFRBOCAAYNHRTD 595  
QY 397 --OEFMGOLYOWEP--FTEVQSGORCELNCRPRGRFFVYRHTKRVQDGTLOQPGAPDICVA 453  
Db 596 LFKSPGPM-DWPRYTGAVPDQCKLTCQARALGYVYVLEPRVVDGTPCSPDSSVVCVQ 654  
QY 454 GCLSPGCGDLGSGRRPDGCGVGGDDSTCLVSGNLTDRGGLGTQKILWIPAGALRL 513  
Db 655 GRCIHAGCDRIIGSKKFKDKCMVCGGSGGCKSGSP--RKFRIYGNVNVVTIPAGATHI 712  
QY 514 QIAOL-RPS--SNYLALRPGGRSITINGNAV--DPPGSRAGGTVFYRNPREECKGE 568  
Db 713 LVROQNGPHRSIYALALPDGSYALNEGYILMPSPTDVLPGLAVSLYS---GATRASE 769  
QY 569 SLSAEGTTPQVDVYMIIF--QENPGVYQYVYISPPPILENPTPEPPVQLOP--EILR 624  
Db 770 TLSGHPGLAQLTLQVLVAGNPQDRLVYSFVPRPTPTPTPQDMLHRRRAQLEILR 829  
QY 625 VEP 527  
Db 830 RRP 832  
RESULT 15  
AT58\_MOUSE  
ID AT58\_MOUSE STANDARD; PRT; 905 AA.  
AC P57110;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE ADAMTS-8 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).  
GN ADAMTS8  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20079168; PubMed=10610729;  
RA Georgiadis K.E., Hirahata S., Seldin M.F., Apte S.S.;  
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";  
RL Genomics 62:312-315(1999).  
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).  
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.  
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AF175282; AAF25805.1; --  
DR MGD; MGI:1353468; Adamts8.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR001590; Reprolysin.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF01421; Reprolysin; 1.  
DR Pfam; PF00090; tspl-1; 2.  
DR SMART; SMO0209; tspl-1; 2.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
DR PROSITE; PS0215; ADAM\_MPRO; 1.  
DR PROSITE; PS00092; TSPI; 1.  
DR PROSITE; PS00427; DISINTEGRIN-1; FALSE\_NEG.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix; Heparin-binding.  
FT SIGNAL 1 28  
FT PROPEP 29 228  
FT CHAIN 229 905  
FT METAL 378 378  
FT ACT\_SITE 379 379  
FT METAL 382 382  
FT METAL 388 388  
FT METAL 453 541  
FT DOMAIN 453 598  
FT DOMAIN 542 598  
FT DOMAIN 599 705  
FT DOMAIN 706 847  
FT DOMAIN 848 905  
FT CARBOHYD 415 415  
FT CARBOHYD 480 480  
FT CARBOHYD 506 506  
FT CARBOHYD 615 615  
SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;  
Query Match 8.2%; Score 399.5; DB 1; Length 905;  
Local Similarity 27.2%; Pred. No. 1.9e-11;  
Matches 123; Conservative 59; Mismatches 161; Indels 109; Gaps 18;  
QY 321 GTPHGR-----LEPDQHP-----GAWLLSNGPHASSLSLWLPAPS----- 358  
Db 514 GTPCGPHCLDGSCLKEDVENPKAVVDGWDGWRP-----WGQCSRTCGGI 562  
QY 359 -----SPIPR-----CSGESEQLRACSQAPCPPEQPDRALQCA---AFNSQEFMGQ 402  
Db 563 QFSNRECDNPNPQNGRFLGKRVKYSCTECPNGKSFQEQCEKYNAYNHTDLGN 622  
QY 403 LYQWEP--FTEVQSGORCELNCRPRGRFFVYRHTKRVQDGTLOQPGAPDICVAGRLSPGC 461  
Db 623 FLQWPYKYSVSPDRCKLFCRARGRSEFKVFEAKVIDGTILCGPDTLSICVYRGQCVKAGC 682  
QY 462 DGIILSGRRPDGCGVGGDDSTCLVSGNLTDRGGLGTQKILWIPAGALRLQIAQ--- 517  
Db 683 DHVNSPKLKDCKGCGGKGTACRKISGFTFFS--GYNDIVIPAGATNIDVKQRSH 740  
QY 518 -LRPSSNYLALRGPGRSIINGNAVDP--PGSYRAGGTVFYRNPREEKGESLSAEGP 575  
Db 741 GVRNDGSYALKTANGQYLLNGNLAIISAIEQDILVKGTILKYSGMATLERLQSFQA--- 797  
QY 576 TQPDVYMIQEENPGVYQYVYISSPPPILENPTPEPPVQLOPEILRVPEPLAPAPRP 635  
Db 798 -----LPEP-----LTVQLTSGVEVFP-PKV 818  
QY 636 ART-----PGLQVRQVRIPQMPAPPHPRTPLGSPAAYKRVKHSACSACSGKGVWPIFCI 692  
Db 819 RYTFVPNDMDFSVQNSKERATTNLIQSL--PSAEWVLGDWSECPSTC-RGSWORRTVEC 875  
QY 693 SRESSEELDERSCAAGAPPPASPECHGTGCP 724  
Db 876 RDPGQASD--TCDE-ALKPEDAKPCGSPQCP 904  
Search completed: July 23, 2002, 21:13:24

us-10-041-770-2.rsp

Wed Jul 24 11:39:35 2002

Job time: 478 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 20:20:06 ; Search time 43.09 Seconds  
(without alignments)  
1955.682 Million cell updates/sec

Title: US-10-041-770-2

Perfect score: 4895

Sequence: 1 MENWGRPWLYLLLSLPQ.....PPAISCILGHAQTSAPPA 877

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	640.5	13.1	1558	2 C89114	protein C37C3.6a [
2	640.5	13.1	2167	2 T34395	hypothetical prote
3	618	12.6	951	2 T00260	hypothetical prote
4	618	12.6	2165	2 T21371	hypothetical prote
5	531	10.8	1205	2 T18517	procollagen N-endo
6	486.5	9.9	951	2 T00017	gene ADAMTS-1 prot
7	466	9.5	550	2 T47158	hypothetical prote
8	414.5	8.5	898	2 T14764	hypothetical prote
9	403.5	8.2	1059	2 T22545	hypothetical prote
10	399.5	8.2	837	2 T00355	hypothetical prote
11	284	5.8	788	2 T25061	hypothetical prote
12	277.5	5.7	2142	2 B35098	MHC class III hist
13	276.5	5.6	1870	2 S37671	MHC class III hist
14	276.5	5.6	1872	2 S36152	MHC class III hist
15	262	5.4	660	1 Q0B63	BHLF1 protein - hu
16	260	5.3	2715	2 T13049	eyelid - fruit fly
17	255	5.2	2649	2 A38587	hypothetical prote
18	251	5.1	1146	2 A38587	collagen, cornea's
19	250.5	5.1	924	2 S27923	gene Irf3 protein -
20	248	5.1	3530	2 A59266	unconventional myo
21	244	5.0	3511	2 A59295	unconventional myo
22	242.5	5.0	1184	2 S50832	atrophin-1 - huma
23	238.5	4.9	963	2 T13140	hypothetical prote
24	237	4.8	957	2 T13976	hypothetical prote
25	235.5	4.8	620	2 S06733	hydroxyproline-ric
26	235.5	4.8	1572	2 T00027	brain-specific ang
27	226.5	4.6	1532	2 A61262	collagen alpha 1(X
28	225.5	4.6	1184	2 G01763	atrophin-1 - huma
29	225.5	4.6	1584	2 T00026	brain-specific ang

30	224	4.6	1151	2 T18535	high molecular mas
31	222.5	4.5	1444	2 T18856	angiogenesis inhib
32	220	4.5	4957	2 T03455	ALR protein - huma
33	220	4.5	5262	2 T03454	ALR protein - huma
34	214.5	4.4	712	2 A45638	immunodominant mic
35	214	4.4	1736	2 T00391	hypothetical prote
36	213.5	4.4	1464	1 CGHU15	collagen alpha 1(I
37	212.5	4.3	1496	1 CGHU2V	collagen alpha 2(V
38	212	4.3	1433	2 A45053	bullous pemphigoid
39	212	4.3	1690	1 CGHU1B	collagen alpha 4(I
40	211	4.3	1546	1 CGHU2E	collagen alpha 1(X
41	210.5	4.3	1049	1 CGBO7S	collagen alpha 1(I
42	210	4.3	1838	1 CGHU1V	collagen alpha 1(X
43	209.5	4.3	1774	2 B56101	collagen alpha 1(X
44	209	4.3	886	2 I50694	collagen alpha 1(I
45	208.5	4.3	753	2 JQ0532	OP protein - Kenne

ALIGNMENTS

RESULT 1

C89114

protein C37C3.6a [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001

C:Accession: C89114

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; PMID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: C89114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1558 <STO>

A:Cross-references: GB:chr\_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CESP:C37C

C:Genetics:

A:Gene: C37C3.6a

A:Map position: 5

Query Match 13.1%; Score 640.5; DB 2; Length 1558;

Best Local Similarity 28.0%; Pred. No. 6.7e-24;

Matches 149; Conservative 77; Mismatches 215; Indels 91; Gaps 12;

Qy 344 NGPHASSIWSLFAFSSPIPR-----CGSESQLRACSOAPCPPEQDPDR 387

Db 71 SGQKETGNWGPWVENECSRCGGVQLKQCSGDCTGASVYISCNLNAC-ESGTDFR 129

Qy 388 ALQCAAFNSQEFMGQLYQWEPFTEVQGSQRCNLNCRGRFRFYVRYHTEKVODGTLQOPGA 447

Db 130 AEQCSKFNDALDGNHKKWTFY---KGKNKCELVCKPESGNFYFKWADKVVYDGTCKDSKS 186

Qy 448 PDICVAGRCISPGDGIILSGRRPDGCGVCGDSTCLVSGNLTDGGPLGYQKILWIP 507

Db 187 NDICVDGCELPVCGDKGSLKFDKCGKCDGSGTKTIEGRFDERNLSPGYHDIKLP 246

Qy 508 AGALRLQIAQLRPSNSYLALRGPGRSRISNGWAVDPGSGYRAGTVFRYRNPREEKG 567

Db 247 EGATWIKIQEARKSTNNLALKNGSDHFLNGLIQVEKEVEVGTFIVY-----DDAEP 301

Qy 568 BLSAEGTTPQVDVYMFQENPGVFQYVYSSPPPILENTPPEPPVQLQPEILRVEP 627

Db 302 ETLQAQGLSEELTALLFRKGRSDTAIKYFSSIP-----LEEEV----- 341

Qy 628 PLAPAPRPARPGTLQROVRIPQMPAPPPTPLGSPAAYWYKRVGHSAACSACSGKGVWRP 687

Db 342 -----DYMFKFDNWTPCSVSCGKGVQTR 364

Qy 688 IFLCISRSGEELDRSCAA-GARPASPECHGTGPPYWEAGETWTSRSCG-PGQHQH 745

Db 425 RVYCHQVANGRRVIVEDGNCITVERPPVKQTCNRFACPEWQAG-PWSACSEKCGDAFOY 483  
QY 806 ROVRC-----VGNNGDEVSEOCASGPPQP-PSREACDMGPTTAMF-HSDWS 851  
Db 484 RSVTCRSEKEGEGKLLAADACPADQEKFDFTERTCNLGPCEGLTFTVTEWNN 535  
RESULT 3  
T00260  
hypothetical protein KIAA0605 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
R:Accession: T00260  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
A:Reference number: Z14086; MUID:98290545  
A:Accession: T00260  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-951 <NAG>  
A:Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAA25531.1; PID:g3043734  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0605  
C:Superfamily: thrombospondin type 1 repeat homology  
P:46-106/Domain: thrombospondin type 1 repeat homology <THRL>  
Query Match 12.6%; Score 618; DB 2; Length 951;  
Best Local Similarity 25.2%; Pred. No. 5 3e-23;  
Matches 192; Conservative 83; Mismatches 235; Indels 252; Gaps 28;  
QY 319 TGGTGHGFRLEPPDPHPGAWLPLLSNGPHASSL-----WSLFA----- 356  
Db 27 TGTSTNDSPTSNS-----LEGGTDATAFWMGWETKWTAFSRSCGGGVTSQERHCL 75  
QY 357 ----PSSPIP---RCSESEQLRACSOAPCPPEQDPRALQCAAFNSQEFMGQLYOWEP- 408  
Db 76 QORRKSVPGNRRTCTGTSKRYQLCRVQECPPDGRSFRQCVSNFVNGRTHQMKPL 135  
QY 409 ----FTEVQSGQRCELACRP-RGFRFVVRHTEKVDGTLCO-POAPDICVAGRLSPGCD 462  
Db 136 YPDYVHI-SSKPCDLHCTTVDGQR---QLMVPARDTSCKLTDLRGVGVSGKEPIGCD 191  
QY 453 GILSGRRPDGCGVGGDDSTCRVLSGNLTDGPGPLCYQKILWIPAGALRIQIACLRPS 522  
Db 192 GVLFTHTLDRKCGICQGGSSCTHTGNYRKNAGHLGYSLVTHIPAGARDIQUIVERKKA 251  
QY 523 NYLALRGPGRSIIINGWAVDPGSRAGTVFRYRNPREEGKG-ESLSAEGTTPQVD 581  
Db 252 DVLALADAGYFFNGYKVDSPKNFNIAIGTVVYRPRMDVYETGIEVIAOGPTNOGLN 311  
QY 582 VYMFQEE-ENPGVYQVVISPPPILENPTPEP-----VP-- 616  
Db 312 V-MVWQNGKSPSITFEYTLQPP---HESRPQIYGFSESASQGLDAGLMGTFPHN 367  
QY 617 ----OLOPEILREPPILAPAP-----RPARTFGTLQR 644  
Db 368 GSLYQASERLGLDNLRFHPGLDMELPGQOQETNEVCQAGGACGPPRGKGRDR 427  
QY 645 QY-----RTPQMPAPHP 657  
Db 428 NYTGTPLTGDKDEVDTHFASQEFSSANAISDQLLGAGSLDKFTLNETVNSIFAQAP 487  
QY 658 RTPLG----- 662  
Db 488 RSLAESFVDEENEGAGPYLLNGSYLELSSDRVANSSSEAPPNVSTSLISAGNRTH 547  
QY 663 -----SPAAY--WKRVGHSAACSCGKGVWRPFLCISRESGELDERSCAA 707  
Db 548 KARTRPKARKQGVSPADMYRWKLSLSHEPCSTCTTGVMSAYAMCV-RYDGVVEDDSYCYDA 606

Db 365 NLYCIDGKNKGRVEDDICEENNAKPKPEFEKSCETVDCAEAWFTGDWESCSSTCGDQGOQY 424  
QY 746 ROLCROQEFGGGSSVPERGCHLPRNITQSCQLRCLGHEVSPWSQSVRCGRGQRS 805  
Db 425 RVYCHQVANGRRVIVEDGNCITVERPPVKQTCNRFACPEWQAG-PWSACSEKCGDAFOY 483  
QY 806 ROVRC-----VGNNGDEVSEOCASGPPQP-PSREACDMGPTTAMF-HSDWS 851  
Db 484 RSVTCRSEKEGEGKLLAADACPADQEKFDFTERTCNLGPCEGLTFTVTEWNN 535  
RESULT 2  
T34395  
hypothetical protein C37C3.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T34395; T34394  
R:Geisler, C.; Bradshaw, H.  
submitted to the EMBL Data Library, July 1996  
A:Description: The sequence of C. elegans cosmid C37C3.  
A:Reference number: Z21518  
A:Accession: T34395  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2167 <GEI>  
A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GNO0023; CESP:C37C3.6b  
A:Experimental source: strain Bristol N2; clone C37C3  
A:Accession: T34394  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1555, 'SKF' <GE2>  
A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GNO0023; CESP:C37C3.6a  
A:Experimental source: strain Bristol N2; clone C37C3  
C:Genetics:  
A:Gene: CESP:C37C3.6b; CESP:C37C3.6a  
A:Map position: 5  
A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/  
Query Match 13.1%; Score 640.5; DB 2; Length 2167;  
Best Local Similarity 28.0%; Pred. No. 9.1e-24;  
Matches 149; Conservative 77; Mismatches 215; Indels 91; Gaps 12;  
QY 344 NGPHASSLSLFASSPIPR-----CSGESQLRACSOAPCPPPQPPDR 387  
Db 71 SGQKETNGPWPVNECSRSGGGVQLEKQCSGDCTGASVRYTSCNLNAC-ESGTDPR 129  
QY 388 ALOCAAFNSQEFMGQLYOWEPFTEVQSGORCELCNCRPRFVVRHTEKVDGTLCPGCA 447  
Db 130 AEQCSKFNDEALDGNHYKWTPT---KGKNCCLVCKPESGNFYKWDKVDGTCDSKS 186  
QY 448 PDCVAGRCILSPGCDGILSGRRPDCGCGVGGDDSTCRVLSGNLTDGPGPLCYQKILWIP 507  
Db 187 NDCVDECLPVGCDKGLSSLKFDKCKGCDGSGTCKTIEGRFDRNLSPGYHDIKLP 246  
QY 508 AGALRIQIQLRPSNYLALRGGRSIIINGWAVDPGSRAGTVFRYRNPREEGKG 567  
Db 247 EGATNRIQIARKSTNLKNGSDHFLNGNLIGLQVEKEVEGGTIFY-----DDAEP 301  
QY 568 ELSAEGTTPQDVYVMIFOENPGVYQVVISPPPILENPTPEPPVQLQPEILRVERP 627  
Db 302 ETLSAQGPLSEETVALLFKGSRDTAIKYEFSIP-----LREEV----- 341  
QY 628 PLAPAPARTPTGLQQRVIRPQMAPPHPTPLGSPAYVKKVYGHSAACSGKGVWRP 687  
Db 342 -----DYMKFEDNWTPCSVSGKGVQTR 364  
QY 688 IFLCISRESGELDERSCAA-GARPPASPEPCHTGPCPPYWEAGETSCSRSG-PGTQH 745  
Db 365 NLYCIDGKNKGRVEDDICEENNAKPKPEFEKSCETVDCAEAWFTGDWESCSSTCGDQGOQY 424  
QY 746 ROLCROQEFGGGSSVPERGCHLPRNITQSCQLRCLGHEVSPWSQSVRCGRGQRS 805

D	b	920	QQSQKLMCLDMSTHROSHDRNCQNVLKPKQAATMCNIDCSTRWITEDVSSCSAKCGSGQK	979			
Q	y	645	QVRIP-----QMPPHH----PRTLPGSPAAY-----WKRVGHSACSASGC-KGV	684			
D	b	980	RQRVCVKMEGDROTASEHLCDRNKSPSDIASCVIDCSGRKNYGEWTSCSETCGSNGK	1039			
Q	y	685	WRPFICISRESGEELDESRCAAGRAPPASPECHGTCPFPYWEAGEWTSRSCSGPGTQ	744			
D	b	1040	MHKRSYCVD-DSNNRRVDESLCGR-EQKATERECNRIPC-FRWVYGHWSECSRCDGGVK	1096			
Q	y	745	HRLQCRQ-----EFG-----GGG-----	758			
D	b	1097	MRHAQLDAADRETHFSRCGPACTQHCHNEHACTTWQFQVMSDCSAKGCDGVQYRDANCT	1156			
Q	y	759	-----SSVPPERCGHLPRPNITOSCQLRGCHWHVEGPSWSQCYSVRGQRQSRYRCVGNN	814			
D	b	1157	DRHRSLVPHRCLKMK-IITPCRHRESPCKYLKE-WSOCVSVEDGWSSRRVSVCSGN	1214			
Q	y	815	GDEVSEQACSGPPSPREACDMPGCTTAWFHSDWSKVSPPEPAISCILGHNAQDT	872			
D	b	1215	GTEVDMSLCGTASDRASHOTCNLGT-C-PFRWNTDWA-----CSVSCGIghrERT	1265			
<hr/>							
RESULT 5							
t18517							
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine							
N;	A;lternate names:	procollagen N-proteinase					
C;	S;pecies:	Bos primigenius taurus (cattle)					
C;	Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999					
C;	Accession:	U18517					
R;	C;ollige, A.; Nusgens, B.V.; Lapiere, C.M.						
A;	Submitted to the EMBL Data Library, February 1996						
A;	Description: Cloning of the cDNA of the bovine procollagen I N-proteinase						
A;	Reference number:	Z18941					
A;	Accession:	U18517					
A;	Status:	Preliminary; translated from GB/EMBL/DDBJ					
A;	Molecule type:	mRNA					
A;	Residues:	1-1205 <COI>					
A;	Cross-references:	EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1					
A;	Experimental source:	skin					
C;	Genetics:						
A;	Gene:	PC I-NP					
C;	Function:						
A;	Description:	catalyzes cleavage of the propeptides of type I and II colla					
C;	Keywords:	hydrolase; metalloproteinase					
<hr/>							
Query Match            10.8%; Score 531; DB 2: Length 1205;							
Best Local Similarity 26.2%; Pred No 1,le-18;							
Matches 164; Conservative 70; Mismatches 235; Indels 156; Gaps							
Q	y	289	WASPQAGRDPDFPSYPGRGOOGQPWGTTGGTPHGKLEPDPOHPGANWLPLLS-----	343			
D	b	546	WLTPDLIKR-----DGNW-----GAWSFPGSCRTC	571			
Q	y	344	-----NGPHASSLWSLFAPSSPTRCSGSEQLRACSAQPCPEQDDPALOCA	392			
D	b	572	GTGVKFTRQCDNP-----PANGRTCSGLAYDQLCNSQCDPALADFEEQCR	622			
Q	y	393	AFNSQ-EFMGOLQWEPFEIVQGSCRELNCRPGRFFRYVHTKPYQDGTLQC-EGARDI	450			
D	b	623	QNDLYFHEDAQHWWLPHEHRDAKERCHLYCESKETGEVVSMKRVMHDGTRSXYKDASF	682			
Q	y	451	CVAGRCLSPGCDGILGSRRPDCGVCGGDDSTCLRVLVSNGLTRDGGPLYQKILWPAGA	510			
D	b	683	CVRGDKRKVGDDIGVISKKOEDKCVCVGDNSSHCHKVTFRSPRKPGLGYIMKFIETPGA	742			
Q	y	511	LRLIOALRPSNYLAURG-PGGRSIINGNWAVDPPG-SYRAGGTVFYRNRPREEGKGE	568			
D	b	743	RHLIIQEADTTHSLAVKNLETGKFFLINEENDVDPSKTFFIAMGVIEWEY---RDGDRE	798			
Q	y	569	SLSAEGPTTOPDVYMIFQ-EENPGVFYQVISSPP-----PILENTPPEPVLOAPE	621			

```
Db 799 TLQTMPLHGTITVLVIGEDARISLTKYKMIHEDSLNVDNNVLED----- 845
QY 622 ILRVEPLAPAPRARTPTGLQVRIPQMPAPPHRTPLGSPAAYKRWKRVGHSACSASCG 681
Db 846 -----DSVGEWALKWSPKPCPG 865
QY 682 KGVNRPFLCISRESGEELDESCAAGARPASPCHGTPC-PPYWEAGBWTSCSRSCG 740
Db 866 GSGQPTKYGCRRLDHKVMHVRGDCSVSKPAIRTCNPQCECQPVVMTGEWEPKSCRSRG 925
QY 741 -PGTHRLQCRQEGGGGS-SVPERCGHLPNNTITQSCOLRLC-GHWEVGSFWSQSV 797
Db 926 RTGMQVRSVRCVQPLHNNITSVTHKHND--ARPEGRACNRELCGRWRAGS-WSQSV 983
QY 798 RCGRQSRQVRC-VGNNGDEVSEQECASGPPQPSREACDMGPC-----TTAWFHS 849
Db 984 TCGNTQERPVLCRTADDSFGVCREE-----RPTARICRLGPCRNTSPSKSYVQ 1037
QY 850 WSKVSPPEPPALCILGHNHAQDTSA 874
Db 1038 WLSRPDPNRP-----VQETSS 1053

RESULT 6
T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.
Genomics 46, 466-471, 1997
A>Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene
A:Reference number: Z14055; MUID:98110583
A:Accession: T00017
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-951 <KUN>
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057
A:Experimental source: strain 129SVJ
C:Genetics:
A:Gene: ADAMTS-1
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2
C:Superfamily: thrombospondin type 1 repeat homology
F:542-598/Domain: thrombospondin type 1 repeat homology <THR>

Query Match 9.9%; Score 486.5; DB 2; Length 951;
Best Local Similarity 23.7%; Pred. No. 1.3e-16;
Matches 149; Conservative 74; Mismatches 186; Indels 219; Gaps 24;

QY 250 RHPRAQAGTEPPPTSLGSGGFFRASQPPRRPSSQGWASQVAGRRPDPFPSPVPRGR 309
Db 508 KUPP--WADGT-----SCGEGKVCVSKCNKTKDKHFATPV-----H 543
QY 310 GOQGO-GPMGT-----GGTPHGRLPEPDQHPGAWLPLLSNGPHASSLSLAFAPSSIP 362
Db 544 GSWGPMGPGDCSRTCGGVQVYTM-RCD-----NEVP 575
QY 363 R-----CSGESQLRACSOAPCPQPDP-RALQCAAFN--SQEPMQ--LYOWEP-FTE 411
Db 576 KNGKYCEGRKRVYASCNTECDPNNGKTFREEQEAHNEFSKASFGNEPVETPKYAG 635
QY 412 VQGSORCELCNRPGRFRFYVHTEKVDGTLQCPGAPDICVAGRCLSPGCGILSGRRP 471
Db 636 VSPKDRCKLTCAKGIYGFVLPQKVPVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFF 695
QY 472 DCGCVGGDDSTCLRVSGNLDRGGPLGQKILWIPAGALRLQIAQ-----LRPSSNYLA 526
Db 696 DKCGVCGGNGSTCKKMSGLVSTRP--GYHDIIVTIPAGATNIEVHRNQRNNGSFLA 753
QY 527 LRGPGRSRIINGNW----AYDPPGSRAGTGVFRYRNPREEKGESLSAE-----GPTTQ 578
Db 754 IRAADGTIYILNGNFTLSTLEQDLTYK--GTVLRS-----GSSAALERIRSFSPKLE 803
```

```
QY 579 P--VDVIMIFQEBNPGVYQYVVISPPPILENPTPEPPVPLQLOPEILRVEPPLAPRPA 636
Db 804 PLTIQVLWGHALRPKIKTYFMKK----- 828
QY 637 RTPGTLQQRVIRPQMPAPPHRTPLGSPAAYKRWKRVGHSACSASCGKGVWRPIPLCISRES 696
Db 829 ----- 828
QY 697 GEELDESCAAGARPASPCHGTPCPCPPYWEAGBWTSCSRSGCGTQHRLOLQCRQEFEG 756
Db 829 -----KTESFNAITPFSWIEBGECSKTCGSGWRVQCRDINGH 871
QY 757 GGSVPPPERGHLPRNITQSCOLRLCGHWEVGSFWSQSVRCGRQSRQVRCVGNNGD 816
Db 872 -----PASECAKEVKPASTRPCADLPCPHQVGD-WSPGSKTCGKGYKKRTLKCVSHDGG 925
QY 817 EVSEQECASGPPQPSR--EACDMGPCT 842
Db 926 VLSNESC--DPLKKPKHYIDFCLTQCS 951

RESULT 7
T47158
hypothetical protein DKFZp762C1110.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (Mewo cell line); clone DKFz762C1110
C:Genetics:
A:Note: DKFZp762C1110.1
```

```
Query Match 9.5%; Score 466; DB 2; Length 550;
Best Local Similarity 22.9%; Pred. No. 7.5e-16;
Matches 136; Conservative 71; Mismatches 179; Indels 208; Gaps 22;

QY 318 GTGG-----TPHGPRLEPDPQHPGAWLPLLSNGP-----HASSLSLFLAP 357
Db 96 GTSGLVLCQTKHFPWADGSCGEGK--CINGKCNKTKDRKHFDTPFHGS--WGMWGP 150
QY 358 -----SSPIPR-----CSGESEQLRACSOAPCPPEQDP-RALQCA 392
Db 151 WDCSRTCGGVQVYTMRECDNPPVNGKGYCEGRVRYRSCNLEDCPDNNGKTFREEQCE 210
QY 393 AFNSQEF-----MGQLYQWEP-FTEVQGSORCELCNRPGRFRFYVHTEKVDGTLQCP 445
Db 211 AHN--EFSKASFGGPAVEWIPKYGVSFDRCKLCOAKGIGYFFVLPQKVVGDGTPCSP 268
QY 446 GAPDICVAGRCLSPGCGILSGRRPDPGCGVGGDDSTCLRVSGNLDRGGLPYQKILW 505
Db 269 DSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKKISGVT--SAKPGYHDIIT 326
QY 506 IPAGALRLQIAQ-----LRPSSNYLALRGPGRSIINGN--AVDPPGSRAGTGVFRY 557
Db 327 IPTGATNIEVHRNQRNNGSFLAKADGTVILNGDYTLSTLEQDIMFK--GVVLR 384
QY 558 NRPREEKGESLSAE-----GPTTQP--VDVYMFQEBNPGVYQYVVISPPPILENPT 610
Db 385 S-----GSSAALERIRSFSPKLEPLTIQVLTGVNLRPKIKTYFMKK----- 427
QY 611 PEPPVPLQLOPEILRVEPPLAPRARTPTGLQVRIPQMPAPPHRTPLGSPAAYWKR 670
Db 428 ----- 427
```

```
QY 671 VCHSACSACGKGVWRPIFLCISRESGEELDERSCAAGARPAPSPBPCHGTPCPYPWEAG 730
Db 428 -----KKESENAITPFSAMWIE 444
QY 731 EWTSCSRGCGTQHOQLCCRCQERFGSSGSSVPPRCGHLPRNITQSCOLRLCGHWEVGS 790
Db 445 EWGECSCGELGWRLVECRDING----QPASECAKEVPASTRCADHPCPQWOLGE 499
QY 791 PWSQSVRCGRQRQRVCGVNGNGDEVSEDEACASGPPPSR--EACDMGPGCT 842
Db 500 -WSSCKTCGKGYKRSKLSKLDHGDVLSHESC--DPLKPKPHIDFCTMAECS 550

RESULT 8
Tl4764
hypothetical protein DKFp434H204.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Tl4764
R:Wanbur, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18161
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-898 <WAM>
A:Cross-references: EMBL:AL110226
A:Experimental source: adult testis; clone DKFp434H204
C:Genetics:
A:Note: DKFp434H204.1

Query Match 8.5%; Score 414.5; DB 2; Length 898;
Best Local Similarity 24.8%; Pred. No. 3.7e-13;
Matches 234; Conservative 63; Mismatches 311; Indels 337; Gaps 53;

QY 19 PQLCDQEVLSGHSLOTPTRECGQEGVGVWQVWASQCGGVQVRSRTC----- 71
Db 133 PRAC---EHLPPPTETPCNRHVPCPATWAVG-NWSQSVTCGEGTQRNVLCTINDTGV 188
QY 72 ---QLPTVOLHPSLPLPPRPHPEALL-PRGQGRPRTSETPLRYTQSRGGGLR 126
Db 189 CDEAQAQPASEVTCISPL---CRWPLGLTLPKSGS-----S 220
QY 127 GPASHLGREETQEI---RAARSRLRDKPKGMFGYGRVFPALPLHNRHRRSPRSEL 183
Db 221 GSSHELFEADFIHPLAPRSPASSP-KPGTMG-----NAIEEAP---EL 264
QY 184 SLISSRGEAIPST-----PRAEPFSANGSPOTELPTLSVHTPSQAERPL 231
Db 265 DL-----PGPVVDFFDYDYNFINHEDLSTGSEEDLDLAGTGDRTTPPHSRPA 315
QY 232 SPETAQTEVAPRTAPALRHHRAQASGTEPPSTHSLGEGGFFRASPQPRPSQGWAS 291
Db 316 APSTGS-----PVP-----ATEPPAAKE---EGVLGPWSPSP-----W-- 345
QY 292 PQVAGRRPDPFSPVGRGQQGQGWGTGGTGHGRLEPDPOH-----PGAWLPLLSNG 345
Db 346 PSQAGRSPPP-PS-----EQ-----TPGNFLINFLPEEDTPIGADPLGLSL-S 388
QY 346 PHASSLSLFAFSPSPTRCSESEQ---LRACSAQPCPEQDPRALCAAFNSQEFMGQ 402
Db 389 PRYST-DGLQTPATP-----ESQNDFFVQKDSQSLPPPPWRDRTN---EVFKDDE--- 434
QY 403 LYONEPTEVQGSQRCELNCRPRGFYVRHTEKVKVDGTLQOPGAPDICVAGRCISPGCD 462
Db 435 -----EPKGRGAPHLPRP-----SSTL-----PP-----LSP--V 458
QY 463 GILGRRPDCGV-CGGDDSTCLRVSGNLTDRGGPLGYOKILW-----IPA 508
Db 459 GSTHSSPSPDVAELWTGGIVAWPEALEGL-----GPVDS--LWPTVGVASLLPPIAPL 512
QY 509 GALLRLQIAQLRPSNLYALRGPGRSIIN-GNMAVDPGSGYRAGGTVFYRNPREEGKG 567
```

```
Db 513 PEMKVRDSSLEPGTSPFPAPGPGSWDLQTVAVWGTFLPTILTGLGHMPALNPGKGP 572
QY 568 ELSAEGTTPQVDVYVMIFQEEENPGVYQYVVISPPPILENPTPEPPVQLOQEILRVEP 627
Db 573 ELSPEVPLS-----SRLST--PAWDSPANSHRVETQP---LAP 608
QY 628 PLAPAPAPARTPGTLQORVRIQMPAPPHPRTPLGSPAAVYKRVGHSACSACGKG-VWR 686
Db 609 SLAEGPPA-----DPLVYNASWQAGNWESECTTTCGLGAVMR 646
QY 687 PIFLCISRESGEELDERSCAAGARP-PASPEPCHGTPCPYPWEAGWTSRSCGPGTQH 745
Db 647 PV-RC---SSGRDED---CAPAGRPOPA--RRCHLRPCAT-WHSGNWSKCSRSCGGSSV 696
QY 746 RLQLOC-----RQFEGGG----- 758
Db 697 RDQVCYDTRDLRLPLRPHPCQPGPAKPPAHPGCAQPCLSWYTSWRECSACGGGQQRL 756
QY 759 -SSVPPRRCGHLPRNITQSCQLRLCGHWEVGPWSQCSVRCGRGORSRQVRCV----- 811
Db 757 VTCPEPGLCEALRPNTTRPCNTHPCQTQWVVG-PWGCSAPCGGVQRRLVKCVNTQIGL 815
QY 812 -----GNGDEVSEDEACAS---GPPQP--SREACDMGPGCTT 843
Db 816 PEEDSDQCGHEAWPESRRCGTEDEPEVPPRCERDLRSFGCET 860

RESULT 9
T22545
hypothetical protein F53B6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22545
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19578
A:Accession: T22545
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1059 <WIL>
A:Cross-references: EMBL:Z81086; PIDN:CA03121.1; GSPDB:GN00019; CESP:F53B6.2
A:Experimental source: clone F53B6
C:Genetics:
A:Gene: CESP:F53B6.2
A:Map position: 1
A:Introns: 38/3; 92/3; 131/3; 169/1; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3;

Query Match 8.2%; Score 403.5; DB 2; Length 1059;
Best Local Similarity 22.9%; Pred. No. 1.5e-12;
Matches 121; Conservative 42; Mismatches 151; Indels 215; Gaps 16;

QY 347 HASSL-WSLFAPSP-----IPRCGESEQLRACSAQPCPPQPDPA 388
Db 20 HVDALSWAASPWSCTNTKCGGVSRLRLCLTSKSGESYRFKVCQKTCESRLRAD 79
QY 389 LQCAAFNSQEFMGQLYQWEPTEVQGSQRCELNCRPR--GFRFVVRHTEKVDQGTLCQPG 446
Db 80 TICGG-----EEIVSRQCEVCSRSLTGANFLWR---VDDGTPCOA 119
QY 447 -APDTCVAGRLSPGCDGILSGRRPDCGVCVGGDDSTCLRVSGNLTDRGGPLGYOKILW 505
Db 120 TSRAVCSKSGCOIVGCDGLISSFRFDACGVCGRGDTG-----DNG-----KFIW 165
QY 506 IPAGALRLQIAQLRPSNLYALRGPGRSIINGNMAVDPGSGYRAGGTVFYRNPREEG 565
Db 166 -----KVSEETACAS-----NCDDIVD----- 183
QY 566 KGESLSAEGTTPQVDVYVMIFQEEENPGVYQYVVISPPPILENPTPEPPVQLOQEILR 625
Db 184 ----- 183
```



QY 626 EPPLAPAPRARTPGTLQROVRIPQMPAPHPRTPLGSPAAWKKRVGHSHACSACGKGW 685  
Db 184 -----WSAGRSIASTS----- 195  
QY 686 RPIFLCISRESGELDERSCAAGARPPASFECHGTGCPYWEAGETWSCSSCGPGTQH 745  
Db 196 QPIVVCVNAITGRVWPEKLCADLRKVEARPCPMLICPSRWMAADWTECVPHCGEGTRK 255  
QY 746 ROI-OCROERGGSSVPPERCCHLHPNITQSCRLCHGHEVGPWQSCVRCRGORS 805  
Db 256 REVICYQTAHVTVHVDFTFCENGTPAAEENCVSTSCGRWEAGK-WSKTASCQGVRR 314  
QY 806 ROYRCVGNNGNDEYSEOCASGPPQPPSRACDMG-PCTTAMFHSWSSK 853  
Db 315 RHVACVGS-----DCDEG-GRPQETTCYAGIPCSTATNSLDWDR 355

RESULT 10  
T00355  
hypothetical protein KIAA0688 - human  
C:Species: Homo sapiens (man)  
C>Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
R:Accession: T00355  
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.  
DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete  
A:Reference number: Z14142; MUID:98403880  
A:Accession: T00355  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-937 <ISH>  
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BA031663.1; PID:g3327190  
A:Experimental source: brain  
C:Genetics:  
A:Gene: KIAA0688  
C:Superfamily: thrombospondin type 1 repeat homology  
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 8.2%; Score 399.5; DB 2: Length 837;  
Best Local Similarity 25.3%; Pred. No. 1,8e-12;  
Matches 168; Conservative 70; Mismatches 264; Indels 151; Gaps 35;

QY 61 GVGVRRTSCQLPTVOL--HPSLPLPPRPHRPHALLPR-----GQGR-PQTSPP---E 109  
Db 235 GAGLKRYLLTVMAAAKAFKPSI-----RNPVSLVTVLVILGSEGGPQVGPSAAQ 287  
QY 110 TLPLYRTQSRGGLRGPASHLG-----RETEQIRAAARSRLRDPKPGMFGYGRVP 163  
Db 288 TLRSCAWQGLNTPEDSDPHEDFTAILTRDLCGVSTC-----DTL--GMADVGV- 338  
QY 164 FALPLHNRHRRSPRSELSISRG--EATPSPTPRAEPTSA---NGSPOTEL-PPTTE 218  
Db 339 -----CDPARSCAIVEDDGLQSAFTAALHGHVFNMLHDSKPCISLNGPLS 385  
QY 219 LSVHTSPQAEPLSPETAQTEVAPTRPAPLRHHPRAQASG-----TEPPSPTH----- 267  
Db 386 TSHVWAPYMAHYDPEPNS-----PCSAFTITFDLNGYCHLLDKPEAPLHPVTF 438  
QY 268 -----SLGEGGFRASQPRRPSQGWASQVAGRRPDPFPSPVRRGQGGGP 316  
Db 439 PGKDYDADROCLTFTGDSRHCPQLPPPCALWCSCHLNG-----HAMCQTKSP 488  
QY 317 WGTGGTTPHGR-----LEPD-----POHPGAWLLLSNGPHASSLSLFPSS-- 359  
Db 489 W-ADGTPCGPAQACMGRLHMDQLDQFNIPQ-AGW-----GP-----WGPWDCST 535  
QY 360 -----PIPR-----CSGESQLRACSOAPCPPEQP--DPRALQCAAFNS--- 396  
Db 536 CGGVQVFSRDCTRPVNRNGKYCEGRRTRFSCNTEDCPTGSAITFREEOCAAYNHRTD 595  
QY 397 --OEEFGQLQWEP-FTEVQGSORCELCNCRGFRFYVRHTEKVODGTLCOPGADICVA 453

Db 596 LKSFQPGM-DWVPRYTVGAPQDQCKLTQCARALGYVYVLEPRVVDGTPCSPDSSVCVQ 654  
QY 454 GRCLSPGCDGILSGRRPDPGCGVCGDDSTCLVSGNLTDRGGPLGYQKILWIPAGALRL 513  
Db 655 GRCHAGCDRIIGSKKFKDKCMVCGDGGSCSKQSGSF--RFRYGYNNVVTIPAGATHI 712  
QY 514 QIAQL-RPS--SNYALALRGPGRSIIINGWAV--DPPGYSRAGGTVERYNRPREGKGE 568  
Db 713 LVYQGNFGRHSIYLAIKLPDGSYALNGEYTLMPSETDVVLPFNAVSLYS---GATAASE 769  
QY 569 SLNAEGTTPQVDVYMF--QENPGVYQYVYISSPPPLENTPPEPPVQLOP--EILR 624  
Db 770 TLSGHGFLAQPLTLQVLVAGNPQDRLRYSFVPRPTPTPRTPQDMLHRAQLILEIR 829  
QY 625 VEP 627  
Db 830 RRP 832

RESULT 11  
T25061  
hypothetical protein T21B6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T25061  
R:Cottage, A.  
Submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19975  
A:Accession: T25061  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-788 <WIL>  
A:Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GNO0028; CESP:T21B6.3  
A:Experimental source: clone T21B6  
C:Genetics:  
A:Gene: CESP:T21B6.3  
A:Map position: X  
A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 5.8%; Score 284; DB 2: Length 788;  
Best Local Similarity 20.2%; Pred. No. 7e-07;  
Matches 188; Conservative 76; Mismatches 296; Indels 370; Gaps 43;

QY 15 LLSLPOLCLDQEVLSGSLQTPTTEG-----QGPEGV--WGPWVQWASC- 56  
Db 115 IVRMLELVDCHQFDAH---PPLEEDGLASSTDKVDKRRQCGDEAVEGVGDWTDWHS 171  
QY 57 -----SQPCGVGVORRSTCOLPTVOLHPSLPLP-----PRPRHPEALLPRGQGR 103  
Db 172 SNHEVRSQACEYGRKIQRGC--PARSAPQVPAPAOQYAPRAPEYPSAQOQQOQREQ 229  
QY 104 PQTSPEPLPLYRTQSRGGLRGPASHLGREETQIRAAARSRLRDPKPGMFGYGRVP 163  
Db 230 QQ-----RQQH-----REHQARLQHQOQQOQQOQQOQQOQQOQPP-----Qp 264  
QY 164 FALPLHNRHRRSPRS-----ELSLSSGEEALPSPTPRAEFPFANGSPQTELPTE 218  
Db 265 QPQFPQPPORPQQQCSFSGTHLHQRCQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 324  
QY 219 LSVHTSPQAEPLSPETAQTEVAPTRP---APLRHHP-----RAQASGTEPPS 264  
Db 325 FGQSIQLOSGPVPPOQHQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 384  
QY 265 PTHSLGEGG--FRASQPRRPSQGWASQVAGRRPDPFPSPVRRGQCGQGWGTGG 321  
Db 385 PKADPCGGCAVPQAPQOQERTPPPPVPLAPINTATQPLFQ----- 427  
QY 322 TPHGRLPDPQHPGAW-----LPLLSNGPHASSLSLFPSSIPRCS-----GESE 369  
Db 428 -PYPTRYRPAAPPACDGGCVNPPVWGVHWDSDWS-----TCSTCTGDGAKS 477  
QY 370 QLRACSOAPCPPEQDPRALQCAAFNSQEFMGOLYQWEPFTEVQGSORCELCNCRGFRF 429

```

Db 478 RRECSTNNC-----OGADYETP-----CNLG-----500
QY 430 YVRHTEKVDQTLQCPGAPDICVAGRCLSGCDGILGSGRRPD-----GCGVCGDDST 483
Db 501 -----PCQWTS-EMCEWSTC-SASC-----GSGQRERTRCHLGTNRCEGKDYE 542
QY 484 CRLVSGNLTDEGGPLGQKILWTPAGALRLQIAQLRPSSNYLALRPGGSRGSIINGNAVD 543
Db 543 SEQCS-----AGPCP-ENSCWEDWQCSTCCQ-----GVAVR-----QRTCLGG-----581
QY 544 PPGSYRAGGVFRYRNPREEGKESLSAEGTTPQVDVYMFQENPGVFYQYVSSPP 603
Db 582 -----VF-----GDHL-CQPKTE-----QRACDGP 602
QY 604 PILENTPPEVPOLQPEILRVEPIAPAPRPAETCTLQVRIRPQMPAPPHPTPLGS 663
Db 603 CSLSPWQE-----611
QY 664 PAAWKRVGHSAASACGKGVWRPIFLCISRESGEELDSRSCAAGA-----RPPASPPCHG 720
Db 612 ---W-----STCSASCGSMKR-----RQVCOFGTDCQGNESQFCYG 648
QY 721 TPCPYWEAEWSCRSRSGFGTQHQLQCRQEFGGGSSVPPERCCHLPRNI-TQSCQ 779
Db 649 PCAEWTEWCEWSCSKCGFGQRTTRGC---LGPNG-----QEATTCQGPSIETLCE 700
QY 780 LRLCGHVEGSPWQSCSVRCGRGQ-----RSRQVRCVG 812
Db 701 GQSCCNWSEWCHWCMDCGCGGVRYEYMTGCEWSPCSQLACEVGVQSRQCVG 760
QY 813 NNG-----DEVSEQ-----ECASGPPPP 831
Db 761 ESGCHCIGLAESQCCGLTQC---PPXPP 787

```

## RESULT 12

MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Aug-1990 #sequence\_revision 06-Nov-1992 #text\_change 24-Aug-2001  
 C:Accession: B35098  
 R:Banerji, J.; Sands, J.; Strominger, J.L.; Spies, T.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990  
 A:Title: A gene pair from the human major histocompatibility complex encodes large protein  
 A:Reference number: A35098; MUID:90192810  
 A:Accession: B35098  
 A>Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2142 <BAN>  
 A:Cross-references: GB:M33509; NID:g179338; PIDN:AAA35585.1; PID:g179339; GB:M31293  
 A:Note: the authors translated the codon AGT for residue 97 as Gly  
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 5.7%; Score 277.5; DB 2; Length 2142;  
 Best Local Similarity 22.3%; Pred. No. 3.6e-06;  
 Matches 226; Conservative 78; Mismatches 359; Indels 351; Gaps 53;

```

QY 24 DQEVLSGH-SLQPTPERGQGE--GYWG-----PWQWASCSQPCGVGVQRRSRTC 71
Db 351 DEGAEGHRDSQASGSEERPEADGKGNSEPEPTPTKTAWETSPP-----398
QY 72 QLPVQVHPSLPLPPRP--RHPEALLPRGCGPRQSPETPLPYRQSRGGRGPIRGPS 130
Db 399 --PTEGPPAPKPLPPGDPYDRGGPPKPPAPDEDEANQRQKQSSEISLAVERAR 456
QY 131 HLGREETQEIARAARRLRDPIKPMFGYGRVPTFALFLHNRHRRHPPRPSRSELSISSRG 190
Db 457 RREERERMQEERRAACAEKLV-----RL 481
QY 191 EEATPSTPR--ABPPFANGSPQTELPPTLSVHTSPQAEPLSPETAQTEVAPRTRAP 248

```

```

Db 482 DEKGAPDKRLKAEPAAAPASTPAPPPAVPKELFAPPAPP--PASAPT---PETEPEE 536
QY 249 LRHHPRAQASGTE--PPSPHSLCEG-----GFRASP-QPRRPSSQGWASQVAGRR 298
Db 537 PAQAPPAQSTPTPGVAAAPTLVSGGSTSTSSGSEASVEPOLSKGPEPEPEV--P 594
QY 299 PDPPSPYPR-----GRGQOQGP-WGTGTGPHGRPLEDDPQHPGAWLPLLSNGPHASSL 351
Db 595 PPTTPPVKVEPKGDGIGTTRQPPSOGLGYPKYQKSLPPRFQROQEQEQLLKQOQHQ---651
QY 352 WSLF-----APSSPIPRCSGESEQLRACSQAPCPPEQDPDRALQCAAFNSQEPMGOLYQWE 407
Db 652 WQHQHQSAPPTPVPPSPQPVTLGAVPAQAPP--PPPKAL-----YFGLGRPP 700
QY 408 PFTVEQSGQCELMCRPRGRFR--YVRHTEKVOGTLCQPGAPDICVAGRCLSPGCDGIL 465
Db 701 PMP-----PMNFDPRMMIPPYV-----DPRLLQ-GRPPL-----EFYPP--GVH 737
QY 466 GSG-----RRPDGCGVCGDDSTCRLVSGNLTDRGGLGYQKILWI-----PA-----508
Db 738 PSGLVPRERSDSLGL--SSEPFDRHAPAMLRERGTTPVDPKLAWGDVFTATPAERPLT 795
QY 509 -----GALRLQIAQLRPSNYLA-----LRGPGGRSIING 538
Db 796 SPLRQAADDDKMRSETPFVPPPPPYLASYPGPENGAPGPPISRFPLEEGPRPL---852
QY 539 NWAYDPPGSYRAGGTVPYRNP--REEKGESLSAEGPTT-----577
Db 853 PW---PQGS---DEVAKIQTPPPKPEKETAQLTGPENGRKLPASRGAGCPPPRE 905
QY 578 -----QPVDYVIMFOENGVFYQYVSSPPFPLENPTPEPVV---QLQPEILR 624
Db 906 SRTETRMGPRGPRSSRRGIPPEEPCA-----PPRRAGPIKKPPPTKVLELPKPLE 956
QY 625 V--EPPLAPAPRPAR-TPGTQLQVRIPQMPAPPHPRTPLG--SPAA-----YMKRVGH 673
Db 957 QGDETTPKPPKPDPLKITGKL-----GGPKETPPNGNLSAPRLRRDYSEYRGP 1006
QY 674 SACSASCGKGVWRPIFLCISR-----ESGEELDSRSCAAGAR---710
Db 1007 TSC-----RGRGGEYFARGGRFYGGRRGGQANSVTFESFEMMGVVEVGGDQTL 1061
QY 711 --PPASPEPCRGTPCPPIWEAGETWTSRSCGPGTQHRQLQCRQEFGGGSSVPPRCGH 768
Db 1062 LLPAAAMPARHGA-----RVQSMRKSFGAGSGAKQAA--1095
QY 769 LPRP-NITQSCQLRLOGHVEGSPWQSCSVRCGCGRSR-----QVRCVGNNGDEV 818
Db 1096 --RPMRVINLLQTKRLPH-----PRRHSRPSRSPTRSLHRAPARFTCPGVGESL 1148
QY 819 SEQECASG-----PQO-----PPSREACDMGCTTAWFHSWSSKVSPEPPA 860
Db 1149 PEGAISPGRRRREAPPQVCPGWSPPAKSLAPKKPPT-----GFLPPS 1190

```

## RESULT 13

S37671  
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 15-Sep-2000  
 C:Accession: S37671  
 R:Bougueleret, L.  
 submitted to the EMBL Data Library, August 1992  
 A:Reference number: S37671  
 A:Accession: S37671  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1870 <BOU>  
 A:Cross-references: EMBL:Z15025; NID:g29374; PID:g29375  
 C:Genetics:  
 A:Map position: 6p21.3  
 A:Introns: 38/2; 97/2; 129/3; 154/1; 203/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;  
 C:Superfamily: collagen alpha 1(IV) chain

```
Query Match      5.6%; Score 276.5; DB 2; Length 1870;
Best Local Similarity 22.7%; Pred. No. 3.6e-06;
Matches 232; Conservative 75; Mismatches 358; Indels 359; Gaps 56;

QY 24 DQEVLSGH-SLQPTTEGGQPE--GWVG-----PWQWASCSQPCGCVGVQRRSRTC 71
Db 350 DEBGAEGHRDSQASGEERPEADGKKGNPSNSEPTPKTANAETSRP----- 397
QY 72 QLTVTOLHPSLPLPPRRHPEALLPRGOGPRPQSPETPLRYTQSRGGRGLRGPASH 131
Db 398 --PETERGPPAPKPLPPPPHGG--PAGNWGPPGDYD-----RGPP--PCKP 438
QY 132 LGREETQETRAARRSLRDIKCMFGYGRVFPALPLHNR-----RHRSPPRSBELS 184
Db 439 PAPERDEAWRQKRSSEI-----SLAVERARRRREERMOEERRAAC 486
QY 185 LISSRGEALPSTPR--AEPFSANGSPOTELPPTLSVHTSPQAEPLSPETAQTEVAP 242
Db 487 EKLKRLDEKFGADKRLKAEPAAPPAAPSTPAPPAVKELPAPPAP--PASAPT--P 541
QY 243 RTAPAPLRHHPRQAASGTE--PPSPHSLGEG-----GFFRASPPRRSSOGWASP 292
Db 542 EKEPEEPAQAQSTPTPCVAAAPTIVSGGSTSTSSGSEFASVPFQPLSKEGPEPP 601
QY 293 QVAGRRPDPSPSVPR-----GRGQOGQP--WGTGTPHGRPLEDPQHPGAWLPLLSNG 345
Db 602 EEV--PPPTTPPVKVEPKGDGIGTPRQPPSQGLGYPKYSKSLPRFRQOQEQLLKQQ 659
QY 346 PHASSLWSLF---APSSPIPCSGESELACSAQPCPEQDPRAILOCFAFNQSEFWMG 401
Db 660 QHQ---WQHQHQSAGPTTPVPPSPQVTLGAVPAPKAPP--PPPKAL-----YFG 705
QY 402 QLYOWEPFTEVOGSQRCNELNCRPRGRF--YVRHTEKVDGTLCPGAPDICVAGRCCLSP 459
Db 706 ALGRPPMP-----PMNFDPRWMMIPYV-----DPRLLQ-----GR--P 738
QY 460 GCD-----GILGSG---RRPDGCGVCGDDSTCRLVSGNLTDRGGPLGVQKILMI----- 506
Db 739 PLDYPPGVPSGLVPRERSDSGL--SSEPTDRHAPAMLRERGTTPVDPKLAWGDVFT 796
QY 507 --PA-----GALRLQIAQLRPSNNYLA-----SLAVERARRRREERMOEERRAAC 487
Db 797 ATPAERPLTSLRQAADDDKMRSETPPVPPPPYLAHYPCFPENGTPGPPISRFPLE 856
QY 529 GPGRSIIINGNVAVDPPGYSRAGTVFRYRNP--REEKGESLSAEGPTT----- 577
Db 857 EPGRPPL---PW---PPGS---DEVAKIQTTPPKPEKPEETAQLTGPEAGRKLPASRS 906
QY 578 -----QPVDVYMIFOENPNGVFYQYVSSPPPILENPTPEPPV-- 616
Db 907 GAGPPPPRESRTETRWGPRGSSRRGIPPEBGA-----PPRRAGPIKKPPPTK 957
QY 617 --OLOPEILRV--EPPLAPAPRAR--TPGTLOQVRIPOMPAPPHPRTPLG--SPAA--- 666
Db 958 VEELPPAPLQGGDETTPKPKDDPLKTKGKL-----GGPKETPPNGMLSPAPRLR 1007
QY 667 ---YKRVGHSACSACGKGVWRPIFLCISR-----ESGELDER 703
Db 1008 RDSYERVGFTSC---RGRGGEYFARGRGRGTGYGGRGAQANSVTFEEMGV 1062
QY 704 SCAGAR-----PPASPECHGTCPYWEAGEWTSRSCSCGCTGHRLOLQCRQFEGGG 758
Db 1063 EVGGDDTLLLPALPARHGA-----RVQSMRKSFSGAG 1098
QY 759 SSVPPERCGLHPRP--NITQSCQLRGLGHMEVSPWQSCSVRQGRGQRSR-----QV 808
Db 1099 SGAQKQAA---RPMRVILQLTRRLPH-----PRRHSPPRSRSTTPSLRHAPARF 1149
QY 809 RCVGNNDDEVSEQACSG-----PPQ-----PPSREACDMGCPCTTAFHSDWSKYSVP 856
Db 1150 TCPGVGESSLEGAISGPRRRREAPPQVQCGWSPPAKSLAPKKPPT-----GP 1197
```

Query Match 5.6%; Score 276.5; DB 2; Length 1872;  
Best Local Similarity 22.7%; Pred. No. 3.6e-06;  
Matches 232; Conservative 75; Mismatches 358; Indels 359; Gaps 56;

QY 24 DQEVLSGH-SLQPTTEGGQPE--GWVG-----PWQWASCSQPCGCVGVQRRSRTC 71  
Db 351 DEBGAEGHRDSQASGEERPEADGKKGNPSNSEPTPKTANAETSRP----- 398  
QY 72 QLTVTOLHPSLPLPPRRHPEALLPRGOGPRPQSPETPLRYTQSRGGRGLRGPASH 131  
Db 399 --PETERGPPAPKPLPPPPHGG--PAGNWGPPGDYD-----RGPP--PCKP 439  
QY 132 LGREETQETRAARRSLRDIKCMFGYGRVFPALPLHNR-----RHRSPPRSBELS 184  
Db 440 PAPERDEAWRQKRSSEI-----SLAVERARRRREERMOEERRAAC 487  
QY 185 LISSRGEALPSTPR--AEPFSANGSPOTELPPTLSVHTSPQAEPLSPETAQTEVAP 242  
Db 488 EKLKRLDEKFGADKRLKAEPAAPPAAPSTPAPPAVKELPAPPAP--PASAPT--P 542  
QY 243 RTAPAPLRHHPRQAASGTE--PPSPHSLGEG-----GFFRASPPRRSSOGWASP 292  
Db 543 EKEPEEPAQAQSTPTPCVAAAPTIVSGGSTSTSSGSEFASVPFQPLSKEGPEPP 602  
QY 293 QVAGRRPDPSPSVPR-----GRGQOGQP--WGTGTPHGRPLEDPQHPGAWLPLLSNG 345  
Db 603 EEV--PPPTTPPVKVEPKGDGIGTPRQPPSQGLGYPKYSKSLPRFRQOQEQLLKQQ 660  
QY 346 PHASSLWSLF---APSSPIPCSGESELACSAQPCPEQDPRAILOCFAFNQSEFWMG 401  
Db 661 QHQ---WQHQHQSAGPTTPVPPSPQVTLGAVPAPKAPP--PPPKAL-----YFG 706  
QY 402 QLYOWEPFTEVOGSQRCNELNCRPRGRF--YVRHTEKVDGTLCPGAPDICVAGRCCLSP 459  
Db 707 ALGRPPMP-----PMNFDPRWMMIPYV-----DPRLLQ-----GR--P 739  
QY 460 GCD-----GILGSG---RRPDGCGVCGDDSTCRLVSGNLTDRGGPLGVQKILMI----- 506  
Db 740 PLDYPPGVPSGLVPRERSDSGL--SSEPTDRHAPAMLRERGTTPVDPKLAWGDVFT 797  
QY 507 --PA-----GALRLQIAQLRPSNNYLA-----SLAVERARRRREERMOEERRAAC 487  
Db 798 ATPAERPLTSLRQAADDDKMRSETPPVPPPPYLAHYPCFPENGTPGPPISRFPLE 857  
QY 529 GPGRSIIINGNVAVDPPGYSRAGTVFRYRNP--REEKGESLSAEGPTT----- 577

MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - hu  
C:Species: Homo sapiens (man)  
C:Date: 06-Jun-1995 #sequence\_revision 17-Nov-1995 #text\_change 15-Sep-2000  
C:Accession: S36152  
R:Irish F.J.M.; Bouguetel, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Ju  
Nature Genet. 3, 137-145, 1993  
A:Title: Dense Alu clustering and a potential new member of the NFkappaB family with  
A:Reference number: S36152; MUID:93272029  
A:Accession: S36152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1872 <IRI>  
A:Cross-references: EMBL:Z15025  
A:Note: in the authors' translation residues 32-34 are shown after residue 4 and, con  
A:Note: the authors translated the codon AAT for residue 1000 as His  
C:Genetics:  
A:Introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1;  
C:Superfamily: collagen alpha 1(IV) chain

Db 858 EPGRPPL---PW-----PPGS-----DEVAKIOTPPPKKEPKEETAQTLGPEAGKRLPASRS 907  
Qy 578 -----QFVDVIMFOENPGVYQYVVISPPPILENPTPEPPV-- 616  
Db 908 GAGPPPPRRSRTETRGPRGSSRRGIPPEBGA-----PPRAGPIKKPPPTK 958  
Qy 617 --QLOPEILAV--EPPLAPAPRAR--TPGTLOQVRIPOMPAPPPHPTPLG--SPAA--- 566  
Db 959 VEELPPPLQGGDETPKPPKPKTKTKGL-----GPKETPENGSLSPAPRLR 1008  
Qy 667 ---YWKRVGHSASCSGKGWWRPIFLCISR-----BSGELDER 703  
Db 1009 RDYSYRVGTSC-----RCGRGEYFARGRGRGTGYGGRGAQANSVTFSEEMCV 1063  
Qy 704 SCAAGAR-----PPASPECHGTPCPYWEAGEWTSRSCPGTQHRLOLQCRQEGGGG 758  
Db 1064 EVGGDGTTLILLFEALPARHGA-----RVQSMRKSPSGAG 1099  
Qy 759 SSVPERCGHLPAP-NITQSCQLRLCGHMEVGSQVRCGRGORSR-----QV 808  
Db 1100 SGAQKQAA---RPMRVILLOTLRPLH-----PREHSPRSRPTTSPTLHAPARF 1150  
Qy 809 RCVGNNDEVSQECASG-----PQO-----PPSREACDMGCPCTAWFHSMDSSKVS 856  
Db 1151 TCPGVGSSLPFGAISPPRRRREAPQVCPGWSPPAKSLAPKKPT-----GP 1198  
Qy 857 EPPA 860  
Db 1199 LPPS 1202

RESULT 15  
Q0B3  
BHLF1 protein - human herpesvirus 4 (strain B95-8)  
C:Species: human herpesvirus 4, Epstein-Barr virus  
C:Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997  
C:Accession: A03742  
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.  
Mol. Biol. Med. 1, 21-43, 1983  
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus  
A:Reference number: A93065; MUID:85035713  
A:Accession: A03742  
A:Molecule type: DNA  
A:Residues: 1-660 <BAN>  
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H  
Nature 310, 207-211, 1984  
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.  
A:Reference number: A03794; MUID:84270667  
A:Contents: annotation; protein coding region  
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52  
C:Superfamily: human herpesvirus 4 BHLF1 protein

Query Match 5.4%; Score 262; DB 1; Length 660;  
Best Local Similarity 22.4%; Pred. No. 6.9e-06;  
Matches 206; Conservative 51; Mismatches 293; Indels 368; Gaps 50;

Qy 35 TPTEGGQGEVGMVQWVQWASCSQPCGVQVRSRRCQLFTVQLHPSLPLPPRPPHPEA 94  
Db 3 TPCQSRGP-----RTTLP-----HCPDCLGAPDQQR 33  
Qy 95 LLPGQGP-----POTSPTPLYTQSRGGGLRGFASHLGHEETQETRAARR 145  
Db 34 RLPPGWQRTAPTQVGLADASDELDQASGARPGGGRV-----AGRGEGTTPAPSRQ 89  
Qy 146 SRLSDPIKGMFGYGVFFALPHNRHRRPPRSRSLSSRGEAIPSPTPRAEP-- 203  
Db 90 SRTGPAEQADHAHNS-----PTGGCSDPQSPRTRQAGYALGEGSAGLSGRGPRPHAF 144  
Qy 204 ---FSAN--GSPQT-----ELPTSLVHTPSQ--AEPLSPETAQTEVAP-----R 243  
Db 145 QVQMSARNPCPTRWRRRSQAQRGHPGQGRPSGTGGRPAAPGCTPAAPGPGGGA 204

Qy 244 TRPAPLRHHRAOASGTEPPSPHSLGEGFFRASP-----QPRRPS-----S 286  
Db 205 AVPSGATPHERGSGPADPFAA-----RLPERQEPRLQDLAAQRCAPGPPPT 255  
Qy 287 QWASPOVAGRRDDPPSPVGRG-----QOGQPNCTGCTPHGRLPDPPQHP 335  
Db 256 RSGAAAOHTHRRPPGCRSARNPGCPTRWRRSQAQRGHPGAGORPSGP--TGGRAAP 314  
Qy 336 GAWPLLSNGSPHASSLSLSPSPIP---RCSGSEQLRACSAQCPPEQDPPRALQCA 392  
Db 315 GAPCTPAAPGPGGAA-----VPSGATPHERGSGPADPFAA---ARLPERQEPRLQ-- 365  
Qy 393 AFNSQEFMGOLYQWEFTVEQSGORCELNCRPRGRFYVHTEKVDGTLQCPGAPDICI 452  
Db 366 -----DLAAQRC-----PAGEPPT 381  
Qy 453 AGRCLS-----PCDGLGRRRPGCGVCGGDDSTCLVSGNLDRGGPLGYKILWI 506  
Db 382 SGAAAOHTHRRPPGCP---RSARNP-GC-----PTWRRSG--AQRGHP-- 420  
Qy 507 PAGALRLQIAQLRPPSNYLALRGP--GGRSINGNWAVDPPGSYRAGGTVFRYRNPREEG 565  
Db 421 PPGAGQ-----RPS-----OPTGGR-----PAAPGAGCT-----PAAPG 449  
Qy 566 KGESLSA--EGPTTQPVDMYMFQENPGVYQYVISPPPILENPTPEPPVQLOPEILR 624  
Db 450 PGGGAAPVSGATPHEP-----ERGS-----PADPAAARLPPRQEPRL-PQDLA 493  
Qy 625 VEPPLAPAPPARTPGLQVRIPOMPAPPHPTPLGSPAAYWKKVGHSAASCSGKGV 584  
Db 494 AAOQCPAGPPTRSGAAAOQTHRRP-----PGCPR-----SARNPGCPR-T 533  
Qy 685 WRPIFLCISRESGEELDERSCAAGARP-----PASPEPCHGTGCPPYWEAGEWTSCSR 737  
Db 534 WR-----RRSGAQRGHPGAGQRPSTGGRPAAP--GAPGTPAAP----- 573  
Qy 738 SCGPG-----TQHRLOLQCRQEGGG-----CSSLVPERCGHLPNITQSCQLRL 782  
Db 574 --GGGGAAPVSGATPH-----PENGSGPADPAAARLPPER---QEPRLQF----- 615  
Qy 783 CGHWEVGSQVSCVRCGRGRQRVRCVGNNGDEVSEQECASGPP-----Q 829  
Db 616 -----DLAAQRCAPGPPPTSRGAAAOQTHRR 642  
Qy 830 PPS--REACDMGCPCTTAW 845  
Db 643 PPGCPRARNPG-CPRTW 659

Search completed: July 23, 2002, 21:06:30  
Job time: 2784 sec

OM of: US-10-041-770-2 to: GenEmbl.\* out\_format : pfs

Date: Jul 23, 2002 10:28 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 compugen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlh

-Q/cgn2\_1/USPTO.spool/US10041770/runat\_22072002.153135.2679/app\_query.fasta\_1.947

-DB=Geneml -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000

-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500

-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPOP=6.000

-FGAPEXT=7.000 -START=1 -MATRIX=blcsun62 -TRANS=human40.cdi

-DELEXT=7.000 -LIST=45 -DOALIGN=200 -OUTPMT=pfs -THR\_MAX=100 -THR\_MIN=0

-ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM=ext -HEAPSIZE=500

-MINLEN=0 -MAXLEN=200000000 -USER=US10041770 -CGNL=1.5038

-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30

-NO\_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-10-041-770-2

Query length: 877

Database: GenEmbl.\*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 2782.860000

score\_list:

Strid	Orig	ZScore	Escore	Len	Documentation
gb_pat:AX342636	+	4769.00	1381.44	4230	AX342636 Sequence 33 from Pat
gb_pat:AL356356	+	3976.50	1139.91	176550	AL356356 Human DNA sequence
gb_hg:AC092479	+	2870.50	769.90	142023	AC092479 Mus musculus clone
gb_hg:AC093317	+	2851.50	762.87	215810	AC093317 Mus musculus clone
gb_hg:AC053497	+	2070.50	597.60	207815	AC053497 Homo sapiens chrom
gb_hg:AC053497	+	2014.50	581.90	207815	AC053497 Homo sapiens chrom
gb_pat:AF217974	+	1231.00	379.63	2070	AF217974 Homo sapiens clone PH
gb_pat:BC008840	+	935.50	294.52	4.8e-08	BC008840 Homo sapiens, hypothe
gb_pat:ES5265	+	807.50	255.40	7.3e-06	ES5265 Novel metalloprotease a
gb_pat:AX128433	+	806.50	254.87	7.8e-06	AX128433 Sequence 2 from Paten
gb_pat:AX128433	+	797.50	254.43	8.3e-06	AX128433 Sequence 15 from Pat
gb_pat:ES5273	+	796.50	254.14	8.6e-06	ES5273 Novel metalloprotease a
gb_pat:AF163762	+	791.50	252.61	1.0e-05	AF163762 Homo sapiens zinc met
gb_pat:AC023772	+	789.50	252.64	1.0e-05	AC023772 Homo sapiens cDNA FLJ
gb_pat:AX327754	+	784.00	250.84	1.3e-05	AX327754 Sequence 10 from Pat
gb_pat:AX327752	+	784.00	250.57	1.4e-05	AX327752 Sequence 8 from Paten
gb_pat:AX319854	+	780.00	250.19	1.4e-05	AX319854 Sequence 18 from Pat
gb_pat:AX327755	+	759.50	243.43	3.4e-05	AX327755 Sequence 11 from Pat
gb_pat:AX111571	+	757.00	243.07	3.6e-05	AX111571 Sequence 1 from Paten
gb_pat:AX327747	+	757.00	243.06	3.6e-05	AX327747 Sequence 3 from Paten
gb_pat:AX338539	+	757.00	242.80	3.7e-05	AX338539 Sequence 1 from Paten
gb_pat:AX327745	+	757.00	242.80	3.7e-05	AX327745 Sequence 1 from Paten
gb_pat:AX338537	+	757.00	242.80	3.7e-05	AX338537 Sequence 1 from Paten
gb_pat:AX319852	+	753.50	241.47	4.4e-05	AX319852 Sequence 16 from Pat
gb_pat:AF314171	+	738.50	236.55	8.2e-05	AF314171 Mus musculus papilin
gb_in:AF078161	+	731.00	230.25	0.0002	AF078161 Manduca sexta lacuna
gb_in:AF205357	+	722.50	229.77	0.0002	AF205357 Drosophila melanogast
gb_pat:AX111573	+	714.50	232.52	0.0001	AX111573 Sequence 3 from Paten
gb_hg:AC014412	+	699.50	218.94	0.0019	AC014412 Drosophila melanog
gb_in:AC009213	+	699.50	218.95	0.0028	AC009213 Drosophila melanog
gb_in:AC003765	+	699.50	207.76	0.0033	AC003765 Drosophila melanog
gb_in:AX319861	+	698.50	226.34	0.0003	AX319861 Sequence 25 from Pat
gb_in:AF043121	+	695.50	223.47	0.0004	AF043121 Haemochus contortus
gb_in:AF011177	+	676.00	219.01	0.0008	AF011177 Homo sapiens mRNA for
gb_pat:HS250725	+	666.00	215.16	0.0013	HS250725 Homo sapiens mRNA for
gb_pat:AX319860	+	636.50	209.08	0.0028	AX319860 Sequence 32 from Pat
gb_pat:AX342635	+	636.50	206.39	0.0039	AX342635 Sequence 24 from Pat
gb_pat:AX342636	+	601.00	200.01	0.0089	AX342636 Sequence 35 from Pat
gb_pat:HS250725	+	598.50	196.87	0.0133	HS250725 Homo sapiens mRNA for
gb_pat:HS250725	+	596.00	196.78	0.0135	HS250725 Homo sapiens mRNA for

seq\_name: gb\_pat:AX342636

seq\_documentation\_block:

LOCUS AX342636

DEFINITION Sequence 33 from Patent WO0198468.

ACCESSION AX342636

VERSION AX342636.1 GI:18152033

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Yue H., Elliott V.S., Gandhi A.R., Lal P., Au-Young J., Tribouley C.M., Deleage A.M., Baughn M.R., Nguyen D.B., Lee E.A., Hafalia A., Khan F.A., Wallia N.K., Yao M.G., Lu D.A., Patterson C., Tang Y.T., Walsh R.T., Azimzal Y., Ramkumar J., Xu Y. and Reddy R. Patent: WO 0198468-A 33 27-DEC-2001; Incyte Genomics, Inc. (US)

JOURNAL

FEATURES

Source 1. 4230

Organization="Homo sapiens" /db\_xref="taxon:9606"

BASE COUNT 810 a 1411 c 1210 g 799 t

ORIGIN

alignment\_scores:

Quality: 4769.00 Length: 857

Ratio: 5.584 Gaps: 0

Percent Similarity: 99.650 Percent Identity: 99.650

alignment\_block:

US-10-041-770-2 x AX342636

Align seg 1/1 to: AX342636 from: 1 to: 4230

1 MetGluAnTrpThrGlyArgProTrpLeuTrpLeuLeuLeuLeuSe 17

228 ATGGAGAACTGGACTGGCAGGCCCTGCTGTATCTGCTGCTCTCTGTC 277

17 rLeuProGlnLeuCysLeuAspGlnGluValLeuSerGlyHisSerLeuG 34

278 CCTCCCTCAGCTCTGCTGTGATCAGAGGTTGTCGCGACACTCTCTTC 327

34 InThrProThrGluGluGlyGlnGlyProGluGlyValTrpGlyProTrp 50

328 AGACACCTACAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 377

51 ValGlnTrpAlaSerCysSerGlnProCysGlyValGlyValGlnArgAr 67

378 GTCCAGTGGGGCTTGTGCTCCAGCCCTGCGGGTGGGGTGGCGACGACG 427

67 gSerArgThrCysGlnLeuProThrValGlnLeuHisProSerLeuProL 84

428 GAGCGGACAGTGCAGCTCCCTACAGTGCAGTGCAGTGCAGTGCAGTGC 477

84 euProProArgProProArgHisProGluAlaLeuLeuProArgGlyGln 100

478 TCCCTCCCG 527

101 GlyProArgProGlnThrSerProGlnThrLeuProLeuTrpArgThrGl 117

528 GGTCCCGACGCCCGACGACTTCCAGAACCCCTCCCTTGTACAGGACACA 577

117 nSerArgGlyArgGlyGlyProLeuArgGlyProAlaSerHisLeuGlyA 134

578	GTCTCGGGGAAGGGGTGGCCCACTTCGAGGTCCCGCTTCCACCTAGGA	527
134	rgGluGluThrGlnGlnIleArgAlaAlaArgSerArgLeuArgAsp	150
628	GAGAGGAGACCCAGGAGATTTCGAGCGGCCAGGAGGTCCCGGCTTCAGAC	677
151	ProIleuSerProGlyMetPheGlyTyrGlyArgValProPheAlaLeuPr	167
678	CCCATCAGCCAGCAATGTTTCGGTTATGGGAGAGTGCCTTTTGCACTTGC	727
167	oLeuHisArgAsnArgArgHisProArgSerProProArgSerSerGluLeuS	184
728	ACTGCACCGGAACCGCAGGCACCTCGGAGCCACCCAGATCTTGAGCTGT	777
184	erLeuIleSerSerArgGlyGluGluAlaIleProSerProThrProArg	200
778	CCCTGATCTCTTAGAGGGGAGAGCCCTATTTCGTCCTCCCTACTCCAGA	827
201	AlaGluProPheSerAlaAsnGlySerProGlnThrGluLeuProProTh	217
828	GCAGAGCCATCTCCGCAACGCGACGCCCAACTGAGTCCCTCCCTCCAC	877
217	rGluLeuSerValHisThrProSerProGlnAlaGluProLeuSerProG	234
878	AGAACTGTCGTCCACACCCCATCCCCCAAGCAGAACCTCTTAAGCCCTG	927
234	luThrAlaGlnThrGluValAlaProArgThrArgProAlaProLeuArg	250
928	AAACTGCTCAGACAGAGGTGGCCCAACAGACCAAGGCTGCGCCCTACGG	977
251	HisHisProArgAlaGlnAlaSerGlyThrGluProProSerProThrHi	267
978	CATCACCCAGAGCCAGGCTCTGGCAGAGAGCCCTCTACCCACGCA	1027
267	sSerLeuGlyGluGlyGlyPhePheArgAlaSerProGlnProArgArgP	284
1028	CTCCTTAGGAAGAGGTGGCTCTTCCTGTCATCCCTCAGCCACGAGGC	1077
284	roSerSerGlnGlyTrpAlaSerProGlnValAlaGlyArgArgProAsp	300
1078	CAGTCTCCAGGTTGGCCAGTCCCCAGGTAGCAGGAGACGCCCTGAT	1127
301	ProPheProSerValProArgGlyArgGlyGlnGlnGlyGlnGlyProTr	317
1128	CCTTTCTCTCGTCCCTCGGGGCCGAGGCCAGCGAGGCCCAAGGGCTTG	1177
317	pglyThrGlyGlyThrProHisGlyProArgLeuGluProAspProGlnH	334
1178	GGGAACGGGGGGACTCTCTCACGGGCCCGCGCTGGAGCCTGACCTCAGC	1227
334	isProGlyAlaTrpLeuProLeuLeuSerAsnGlyProHisAlaSerSer	350
1228	ACCGGGGGCCCTGGCTGCCCTGCTGAGCAACGCCCCCATGCCAGCTCC	1277
351	LeuTrpSerLeuPheAlaProSerSerProIleProArgCysSerGlyGI	367
1278	CTCTGGAGCCCTTGTGCTCCAGTAGCCCTATTCCAAAGATGTTCTGGGA	1327
367	uSerGluGlnLeuArgAlaCysSerSerGlnAlaProCysProProGluGlnP	384
1328	GAGTGACAGCTAGAGGCTTCAGCCAGCCAGCCCTGCCCTCTGAGGACGC	1377
384	roAspProArgAlaLeuGlnCysAlaAlaPheAsnSerGlnGluPheMet	400
1378	CAGACCCCGGGCCCTGCAGTGGCGAGGCTTTAACTCCAGGAATTCATG	1427
401	GlyGlnLeuTyrGlnTrpGluProPheThrGluValGlnGlySerGlnAr	417
1428	GGCCAGCTGTATCAGTGGAGGCCCTTCACTGAAGTCCAGGGCTCCCCAGC	1477
417	gcysGluLeuAsnCysArgProArgGlyPheArgPheTyrValArgHisT	434

750 erCysSerArgSerCysGlyProGlyThrGlnHisArgGlnLeuGlnCys 750  
|||||  
2428 CCTGACCGCTCTCTGTGGCCCGGACCCAGCACCAGCCAGCTGCAGTC 2477  
751 ArgGlnGluPheGlyGlyGlySerSerValProGluArgCysG1 767  
|||||  
2478 CGCAGGAATTTGGGGGGGTGCTCTCGGTGCCCCCGGAGCGCTGTGG 2527  
767 yHisLeuProArgProAsnIleThrGlnSerCysGlnLeuArgLeuCysG 784  
|||||  
2528 ACATCTCCCCCGGCCCAACATCACCCAGCTTGCCAGCTGCGCCTCTGG 2577  
784 lyHisIrpGluValGlySerProIrpSerGlnCysSerValArgCysGly 800  
|||||  
2578 GCCATTGGGAAGTTGCTCTCTCTGGAGCCAGTGCTCCGTGGGTGCGGC 2627  
801 ArgGlyGlnArgSerArgGlnValArgCysValGlyAsnAsnGlyAspG1 817  
|||||  
2628 CGGGGCCAGAGAAGCGGCGAGGTTCGCTGTGTGGGAACAACGGTGATGA 2677  
817 uValSerGluGlnGluCysAlaSerGlyProGlnProProSerArg 834  
|||||  
2678 AGTGACGAGCAGGAGTGTGCTGACGCCCCCAGCCAGCCAGCAGAG 2727  
834 luAlaCysAspMetGlyProCysThrThrAlaIrpPheHisSerAspIrp 850  
|||||  
2728 AGGCCTGTGACATGGGGCCCTGTACTACTGCTGTTCACAGCGACTGG 2777  
851 SerSerLysValSerProGlu 857  
|||||  
2778 AGCTCCAGTGTCTCAGCCGAG 2798